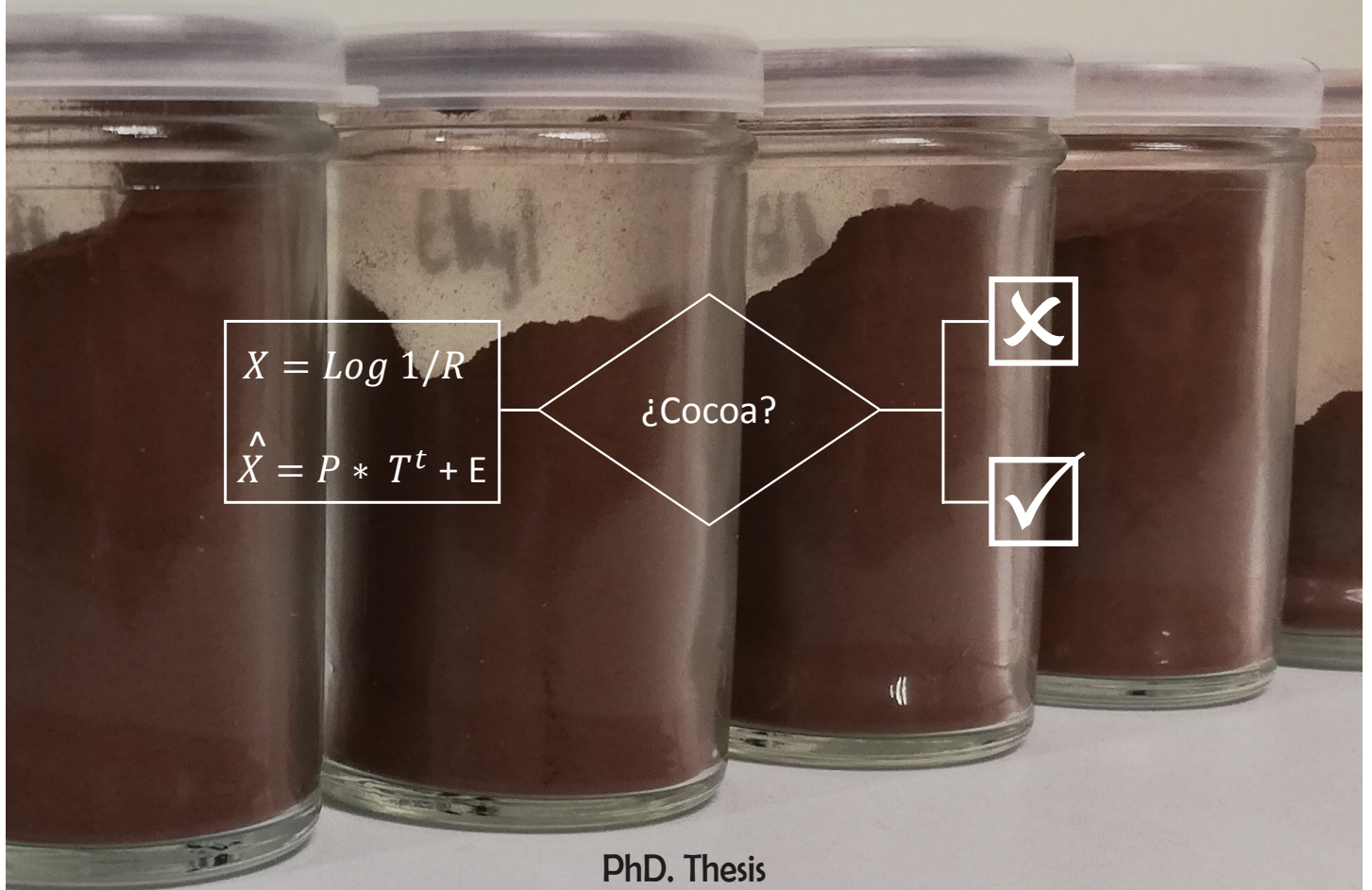UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Application of Infrared Spectroscopy and Chemometrics to the Cocoa Industry for Fast Composition Analysis and Fraud Detection

$$X = Log\ 1/R$$

$$\hat{X} = P * T^t + E$$

¿Cocoa?

✘

✔

PhD. Thesis

Author
Maribel Alexandra Quelal Vásconez

Supervisors
Full. Prof. Pau Talens Oliag · PhD. Édgar Pérez Esteve · PhD. María Jesús Lerma García

Valencia 2019

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

DEPARTAMENTO DE TECNOLOGÍA DE ALIMENTOS

Application of Infrared Spectroscopy and Chemometrics to the Cocoa Industry for Fast Composition Analysis and Fraud Detection

PhD THESIS

Submitted by

Maribel Alexandra Quelal Vásconez

Supervisors:

Full. Prof. Pau Talens Oliag
PhD. Édgar Pérez Esteve
PhD. María Jesús Lerma García

Valencia, December 13, 2019

PAU TALENS OLIAG, PhD in Food Technology and Full Professor at the *Universitat Politècnica de València*, MARÍA JESÚS LERMA GARCÍA PhD in Chemistry and Professor at the *Universitat de València* and ÉDGAR PÉREZ ESTEVE, PhD in Food Technology and Professor at the *Universitat Politècnica de València*

CERTIFY:

That the work "Application of Infrared Spectroscopy and Chemometrics to the Cocoa Industry for Fast Composition Analysis and Fraud Detection" has been developed by Maribel Alexandra Quelal Vásconez under their supervision in the CUINA of the *Universitat Politècnica de València*, as a Thesis Project in order to obtain the degree of PhD in Food Science, Technology and Management at the *Universitat Politècnica de València*.

Valencia, December 13, 2019.

Full.Prof. Pau Talens Oliag

PhD. María Jesús Lerma García   PhD. Édgar Pérez Esteve

"Live as if you were to die tomorrow,
learn as if you were to live forever."
Mahatma Gandhi

Acknowledgements

Agradecimientos

En un mundo lleno de algoritmos, patrones y sesgos, fuerza el cambio, promueve la inexactidud de los modelos establecidos, diviértete siendo un outlier.
La frase mencionada podría resumir este período y parte del trabajo que he llevado a cabo, en parte proviene de la gran admiración que tengo por todo lo que el ser humano puede llegar a hacer, por los innumerables desafíos que puede enfrentar y superar. La Ciencia es una buena oportunidad para nunca dejar de asombrarnos, y me siento muy agradecida por esto.

Resumen

El cacao es un producto de alto valor, no únicamente por sus características sensoriales, sino porque también presenta un alto contenido en antioxidantes y alcaloides estimulantes con efectos saludables. Debido a la alta demanda, la industria del cacao en polvo tiene el desafío de asegurar la calidad de grandes volúmenes de producción de una manera rápida y precisa, evitando la presencia de contaminantes o adulterantes en la materia prima, ofreciendo productos donde se preserven las propiedades saludables. La espectroscopia del infrarrojo cercano (NIR) es una tecnología rápida y no destructiva útil en el análisis de productos alimentarios. La presente tesis doctoral se centra en evaluar el potencial uso del NIR como una herramienta de control de calidad con el fin de poder resolver problemas que se presentan en la industria del cacao en polvo. Los problemas a resolver incluyen la detección de materiales no deseados o adulterantes en el cacao en polvo, y la monitorización rápida y precisa del contenido de flavanoles y metilxantinas del cacao en polvo durante el proceso de alcalinización. El primer capítulo evalúa la viabilidad del NIR, en combinación con análisis quimiométricos, en la detección de la presencia de materiales no deseados o adulterantes como son cascarilla de cacao o harina de algarroba. Para ello, diferentes muestras de cacao en polvo natural y con diferentes niveles de alcalinización (suave, medio y fuerte) fueron mezcladas con distintas proporciones de cascarilla de cacao (con cacao natural) o harina de algarroba (con cacao natural y alcalinizado).

Los resultados obtenidos indican que el NIR, combinado con modelos estadísticos tales como el análisis discriminante por mínimos cuadrados parciales (PLS-DA) y la regresión parcial de mínimos cuadrados (PLS), es un método rápido y eficaz para identificar cualitativa y cuantitativamente materiales no deseados o adulterantes como la cascarilla y la algarroba en cacao en polvo, independientemente del grado de alcalinización o el nivel de tostado de la harina de algarroba. Mediante

Resumen

PLS-DA, para el estudio de la mezcla de polvo de cacao con cascarilla de cacao, el 92,5% de las muestras se clasificaron correctamente en dos grupos: muestras con un contenido de cascarilla inferior al 5% (considerado el límite de aceptación en polvo de cacao por el Codex Alimentarius) y contenidos de cascarilla entre 5 y 40%. En el estudio de la adulteración de cacao en polvo con harina de algarroba, el 100% de las muestras se clasificaron correctamente en tres grupos: cacao, harina de algarroba y mezclas. En ambos estudios, fue posible calcular el porcentaje de adulterante y material no deseado presente en las muestras mediante la construcción de modelos de regresión PLS. En relación a la cascarilla de cacao, el mejor modelo de predicción PLS se obtuvo con 1 LV con un R^2 de 0,975 y 0,967, respectivamente, y un error cuadrático medio de 1,91 y 2,43%, para los conjuntos de calibración y validación externa, respectivamente. Para el estudio de adulteración con harina de algarroba, el modelo de regresión PLS se obtuvo con 1 variable latente (LV) con un R^2 de 0,980 y 0,974, y un error cuadrático medio de 2,9 y 3,2% para los conjuntos de calibración y validación externa, respectivamente.

En el segundo capítulo, el análisis composicional del cacao en polvo se orientó al control de los cambios producidos en el contenido de flavanoles y metilxantinas debidos al proceso de alcalinización al que se somete el caco en polvo. Se determinó el contenido de catequina, epicatequina, cafeína y teobromina mediante cromatografía líquida de alta resolución (HPLC), correlacionándose los contenidos obtenidos para cada uno de estos compuestos con las determinaciones NIR. Se obtuvieron buenos modelos para la predicción de los compuestos mediante regresión PLS con valores superiores a 3 para la relación entre el rendimiento y la desviación (RDP), lo cual demuestra que los modelos obtenidos pueden ser utilizados para la rápida y fiable predicción del contenido de flavanoles y metilxantinas en cacaos naturales y con diferentes niveles de alcalinización.

Resum

El cacau és un producte d'alt valor, no sols per les seues característiques sensorials, sinó perquè també presenta un elevat contingut en antioxidants i alcaloids estimulants amb efectes saludables. A conseqüència a l'alta demanda, l'industria del cacau en pols té el desafiament d'assegurar la qualitat de grans volums de producció d'una manera ràpida i precisa, evitant la presència de contaminants o adulterants en la matèria cosina, oferint productes a on se preserven les propietats saludables. L'espectroscòpia de l'infrarroig proper (NIR) és una tecnologia ràpida i no destructiva útil en l'anàlisi de productes alimentaris. La present tesis doctoral se centra en avaluar el potencial ús del NIR com una eina de control de qualitat amb l'objectiu de poder resoldre problemes que es presenten en l'industria del cacau en pols. Els problemes a resoldre inclouen la detecció de materials no desitjats o adulterants en el cacau en pols, i la monitorització ràpida i precisa del contingut de flavanols i metilxantines del cacau en pols durant el procés d'alcalinització. El primer capítol avalua la viabilitat del NIR, en combinació amb anàlisis quimiométrics, en la detecció de la presència de materials no desitjats o adulterants com són pellofa de cacau o farina de garrofa. Per a això, diferents mostres de cacau en pols natural i amb diferents nivells d'alcalinització (suau, mig i fort) foren barrejades en distintes proporcions de pellofa de cacau (en cacau natural) o farina de garrofa (en cacau natural i alcalinisat).

Els resultats obtinguts per a NIR, combinats amb models estadístics com l'anàlisi discriminant per mínims quadrats parcials (PLS-DA) i la regressió parcial de mínims quadrats (PLS), és un mètode ràpid i eficaç per identificar materials no desitjats o adulterants com la pellofa de cacau o la farina de garrofa, amb independència del grau d'alcalinització del cacau o de torrat de la farina de garrofa. Mitjançant PLS-DA, per a l'estudi de la barreja de pols de cacau amb pellofa de cacau, el 92,5% de les mostres es classifiquen correctament en dos grups: mostres

Resum

amb un contingut de pellofa inferior al 5% (considerat el límit de acceptació en pols de cacau pel Codex Alimentarius) i continguts de pellofa entre 5 i 40%. A l'estudi d'adulteració de cacau en pols amb farina de garrofa, el 100% de les mostres es classifiquen correctament en tres grups: cacau, farina de garrofa i mescles. En ambdós estudis, s'ha pogut calcular el percentatge d'adulterant o material no desitjat present a les mostres mitjançant la construcció de models de regressió PLS. En relació amb la pellofa de cacau, el millor model de predicció PLS s'ha obtingut amb 1 variable latent (LV), amb R^2 de 0,975 i 0,967, i un error quadràtic medi d'1,91 i 2,43%, per als conjunts de calibratge i validació externa, respectivament. Per a l'estudi d'adulteració amb farina de garrofa, el model de regressió PLS s'obté amb una variable latent (LV) amb un R^2 de 0.980 i 0.974, i un error correcte mitjà de 2,9 i 3,2% per als conjunts de calibració i validació externa, respectivament.

En el segon capítol, l'anàlisi composicional del cacau en pols s'orientà al control dels canvis produïts en el contingut de flavanols i metilxantines a causa del procés d'alcalinització al que se sotmet el cacau en pols. Es va determinar el contingut de catequina, epicatequina, cafeïna i teobromina mitjançant cromatografia líquida d'alta resolució (HPLC), i es van correlacionar els continguts obtinguts per a cadascun d'estos composts amb les determinacions NIR. Es van obtindré bons models per a la predicció dels composts mitjançant regressió PLS amb valors superiors a 3 per a la relació entre el rendiment i la desviació (RDP), la qual cosa demostra que els models obtinguts poden ser emprats per a la ràpida i fiable predicció del contingut de flavanols i metilxantines en cacaus naturals o amb diferents nivells d'alcalinització.

Abstract

Cocoa is a product of high value, not only because of its sensory characteristics, but also because it has a high content of antioxidants and stimulating alkaloids with health effects. Due to the high demand, the cocoa powder industry has the challenge of ensuring the quality of large volumes of production in a fast and accurate way, avoiding the presence of contaminants or adulterants in the raw material, offering products where the healthy properties are preserved. The near infrared spectroscopy (NIR) is a rapid and non-destructive technology useful in the analysis of food products. The present doctoral thesis focuses on evaluating the potential use of NIR as a quality control tool in order to solve problems that arise in the cocoa industry powdered. The problems to solve include the detection of unwanted materials or adulterants in the cocoa powder, and the rapid and accurate monitorization of the flavanols and methylxanthines content in the cocoa powder during the alkalization process. The first chapter evaluates the viability of the NIR, in combination with chemometric analysis, in the detection of presence of unwanted materials or adulterants such as cocoa shell or carob flour. For this, different samples of natural cocoa powder and with different levels of alkalization (light, medium and strong) were mixed with different proportions of cocoa shell (with natural cocoa) or carob flour (with natural and alkalized cocoa).

The results obtained indicate that the NIR combined with statistical models such as the partial least squares discriminant analysis (PLS-DA) and the partial least squares regression (PLS), is a fast and efficient method to identify qualitative and quantitative unwanted materials or adulterants such as shell and carob in cocoa powder, regardless of the degree of alkalization or level of roasting of carob flour. By PLS-DA analysis, for the study of the adulteration with cocoa shell, 92.5% of the samples were correctly classified into two groups: samples with a shell content of less than 5% (considered the acceptance limit in cocoa powder by the Codex

Abstract

Alimentarius) and shell contents between 5 and 40%. In the study of the adulteration of cocoa powder with carob flour, the 100% of the samples were correctly classified into three groups: cocoa, carob flour and mixtures. In both studies, was possible to calculate the percentage of adulterant and non-wanted material present in the samples by means of the construction of PLS regression models. In relation to cocoa shell, the best PLS prediction model was obtained with 1 LV, with an R^2 of 0.975 and 0.967, respectively, and an average square error of 1.91 and 2.43%, respectively. For the study of adulteration with carob flour, the PLS model was obtained with 1 latent variable (LV), with an R^2 of 0.980 and 0.974, and a root mean square error (RMSE) of 2.9 and 3.2% for the calibration and external validation sets, respectively.

In the second chapter, the compositional analysis of cocoa powder was oriented to the control of the changes produced in the content of flavanols and methylxanthines due to the process of alkalization to which the cocoa powder is subjected. The content of catechin, epicatechin, caffeine and theobromine were determined by high performance liquid chromatography (HPLC), correlating the contents obtained for each of these compounds with the NIR determinations. Good models were obtained for the prediction of compounds by regression PLS with values above 3 for the ratio of performance to deviation (RDP), which shows that the obtained models can be used for the quick and reliable prediction of flavanol content and methylxanthines in natural cocoas and with different alkalization levels.

PREFACE

This thesis forms part of the project "Estudio de la relación entre variables de procesado y cambios en la composición nutricional y perfil funcional del cacao en polvo. Desarrollo de una metodología predictiva aplicada al procesamiento (RTC-2016-5241-2)", funded by the MINISTERIO DE ECONOMIA Y EMPRESA - AGENCIA ESTATAL DE INVESTIGACION.

The general objective of this project is to develop a predictive methodology for the process variables that lead to a specific final product from a specific raw material.

For this purpose, the relationship between the processing variables and the changes in the nutritional composition and functional profile of cocoa powder might be studied.

To achieve this general objective different specific objectives have been proposed, among which the objective of this thesis is included. This is "to develop rapid methods of analysis that allow to predict the origin of a sample and its behavior during an alkalization process from optical measurements".

The thesis is structured in six sections: introduction, objectives, scientific contribution, general discussion, conclusions and future perspectives. The introduction section focuses on a review of the conventional and alternative methods used for the quality and authenticity control in the cocoa industry. The objectives section presents the general and specific objectives of the thesis. The scientific contribution section is divided in two chapters, the chapter 1, *Identification of cocoa components out of normative limits or non-declared ingredients in cocoa powders,*

and the chapter 2, *Prediction of the content of bioactive compounds of cocoa powders.*

The chapter 1 is devoted to the application of NIR spectroscopy and chemometric algorithms to the quantitative and qualitative determination of cocoa shell and carob powder in cocoa powders.

The results included in this chapter have been published into 2 articles:

- *ARTICLE I. Fast detection of cocoa shell in cocoa powders by near infrared spectroscopy and multivariate analysis.*

- *ARTICLE II. Rapid fraud detection of cocoa powder with carob flour using near infrared spectroscopy*

In chapter 2, the content of some of the main bioactive compounds of cocoa powders, which are flavanols (catechin and epicatechin) and methylxanthines (caffeine and theobromine) in natural and alkalized cocoa powders was established by high-performance liquid chromatography in order to study the influence of cocoa alkalization in their content. Next, these contents were predicted by means of NIR data and chemometrics.

The results included in this chapter are now considered for their publication in 1 article:

- *Article III. Changes in methylxanthines and flavanols during cocoa powder processing and its quantification by near-infrared spectroscopy.*

The general discussion, conclusions, and future perspectives sections presents a short general discussion proposal for further possible studies.

Dissemination of results

International Journals indexed in JCR

Published

Quelal-Vásconez, M. A., Pérez-Esteve, É., Arnau-Bonachera, A., Barat, J. M., & Talens, P. (2018). Rapid fraud detection of cocoa powder with carob flour using near infrared spectroscopy. *Food Control*, 92, 183–189.

Quelal-Vásconez, M. A., Lerma-García, M.J., Pérez-Esteve, É., Arnau-Bonachera, A., Barat, J. M., & Talens, P. (2019). Fast detection of cocoa shell in cocoa powders by near infrared and multivariate analysis. *Food Control*, 99, 62–72.

Quelal-Vásconez, M. A., Lerma-García, M.J., Pérez-Esteve, É., Arnau-Bonachera, A., Barat, J. M., & Talens, P. (2020). Changes in methylxanthines and flavanols during cocoa powder processing and their quantification by near-infrared spectroscopy. *LWT -Food Science and Technology, 117*,108598.

Quelal-Vásconez, M. A., Lerma-García, M.J., Pérez-Esteve, É., Talens, P., & Barat, J. M. (2020). Roadmap of cocoa quality and authenticity control in the industry: a review of conventional and alternative methods. Submitted to Comprehensive Reviews in Food Science and Food Safety.

National and International Conferences

Poster

Quelal-Vásconez, M. A., Torres Mesa, C., Cortes López, V., Lerma García, M. J., Barat, J. M., & Talens, P. (2016). Rapid evaluation of simple sugars in fruit juices and nectars using Fourier-transform infrared spectroscopy and multivariate analysis.

Dissemination of Results

4th International ISEKI_Food Conference (ISEKI_FOOD 2016) Workshop on Sensors and Molecular Recognition, July 6-8.

Quelal-Vásconez, M. A., Pérez-Esteve, É., Arnau-Bonachera, A., Barat, J. M., & Talens, P. (2017). Qualitative and quantitative determination of natural cocoa powder mixed with carob flour by NIR spectroscopy. 3th International and 4th National Student Congress of Food Science and Technology, February 23-24.

Quelal-Vásconez, M. A., Lerma García, M. J., Pérez-Esteve, É., Arnau-Bonachera, A., Barat, J. M., & Talens, P. (2017). Determinación rápida del contenido de alcaloides y polifenoles de cacao en polvo por espectroscopia de infrarrojos. XI International Workshop on Sensors and Molecular Recognition 6-7 july.

Quelal-Vásconez, M. A., Lerma García, M. J., Pérez-Esteve, É., Arnau-Bonachera, A., Barat, J. M., & Talens, P. (2018). Feasibility of Near Infrared Spectroscopy for the detection of cocoa shell in cocoa powders. 4th International and 5th National Student Congress of Food Science and Technology, February 23-24.

Quelal-Vásconez, M. A., Lerma García, M. J., Pérez-Esteve, É., Arnau-Bonachera, A., Barat, J. M., & Talens, P. (2018). Determination of methylxanthines and flavanols in cocoa powders with different alkalization degrees by high performance liquid chromatography. 4th International and 5th National Student Congress of Food Science and Technology, February 23-24.

Scholarship recipient

Doctoral scholarship, National Secretary of Higher Education, Science, Technology and Innovation (Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación – SENESCYT).

Pre-doctoral Stage at a Foreign Institution

Copenhagen University, Department of Food Science Quality and Analytic Technology, Denmark. From 4 Feb to 9 June 2018, under the supervision of Prof. Rasmus Bro.

Abbreviations and Acronyms

AAS	Atomic Absorption Spectrometry
ABTS	2,2′-azino-bis (3-ethylbenzothiazoline-6-sulfonic acid)
AOAC	Association of Official Agricultural Chemists
ANOVA	Analysis of Variance
ACO-PLS	Ant Colony Optimization-PLS
AEDA	Aroma Extraction and Dilution Analysis
ASV	Anodic Stripping Voltammetry
BPANN	Backpropagation Neural Networks
CARS-PLS	Competitive Adaptive Reweighted Sampling - Partial Least Squares
CBE	Cocoa Butter Equivalent
Cd	Cadmiun
CFU	Colony Forming Unit
CV	Cross-Validation
CVA	Canonical Varieties Analysis
DCF	Dark Carob Flour
DFA	Discriminant Funtion Analysis
DM	Dry Matter
DMP	Dimethylpyrazines
DPPH	-2,2-diphenyl-1picrylhydrazil
EFSA	European Food Safety Authority
EMSC	Extended Multiple Signal Correction
ET	Electronic Tongue
ESI	Electrospray Ionization
FAME	Fatty acids methyl esters
FC	Fermented Cocoa
FFA	Free fatty acids
FID	Flame Ionization Detector
FI-ESI-MS	Flow Infusion-Electrospray Ionization-Mass Spectrometry

Abbreviations and Acronyms

FRAP	Ferric Reducing Antioxidant Power
FT	Fourier Transform
FT-IR	Fourier Transform Infrared
FT-NIR	Fourier Transform Near-Infrared
GC-O-MS	Gas Chromatography Olfactometry Mass Spectrometry
HA	High adulteration
HCA	Hierarchical Cluster Analysis
HPLC-MSESI-QTOF	High-performance liquid chromatography coupled to electrospray ionisation and quadrupole time-of-flight mass spectrometry
HSI	Hyperspectral image
HS-SPME	Headspace-Solid Phase Microextraction
IAC	Immunoaffinity Column
ICCO	International Cocoa Organization
ICP-OES	Inductively coupled plasma optical emission spectrometry
ILR-CIS-GC–MS	In-line roasting hyphenated with a cooled injection system coupled to a gas chromatograph–mass spectrometer
InGaAs	Indium gallium arsenide
KNN	Kernel Nearest Neighbor
LA	Low adulteration
LAC	Light Alkalized Cocoa
LCF	Light Carob Flour
LDA	Linear Discriminant Analysis
LHS	Latin Hypercube Strategy
LOO-CV	Leave one cross-validation
LOD	Instrumental Limit of Detection
LOQ	Limits of quantification
LV	Latent Variable
MA	Medium adulteration
MAC	Medium Alkalized Cocoa
MCF	Medium Carob Flour

Abbreviations and Acronyms

MIR	Mid-Infrared
mPLS	Modified Partial Least Squares
MSC	Multiple Scatter Correction
MS	Mass Spectrometry
NC	Natural Cocoa
NER	Non Error Rate
OPA	O-phthalaldehyde
ORAC	Oxygen Radical Antioxidant Capacity
OSC	Orthogonal Signal Correction
OTA	Ochratoxin A
PAHs	Polycyclic Aromatic Hydrocarbons
PARAFAC	Parallel Factor Analysis
PCA	Principal Component Analysis
PDA	Photodiode Array Detector
PLS	Partial least Squares
PLS-DA	Partial Least Squares Discriminant Analysis
QDA	Quadratic Discriminant Analysis
R^2	Coefficient of determination
Rc	Correlation coefficient of calibration
RP-HPLC-DA	Reverse Phase- High Liquid Chromatography Coupled to a Photodiode Array Detector
RMSEC	Root Mean Square Error of Calibration
RMSECV	Root Mean Square Error of Cross-validation
RMSECV	Root Mean Square Error of Cross-Validation
RMSEP	Root Mean Square Error of Prediction
RPD	Ratio prediction deviation
Rp	Correlation coefficient of prediction
RSD	Relative Standard deviation
SAFE	Solvent-Assisted Flavor Evaporation
S-G	Savitzky Golay
SIMCA	Soft Independent Modelling of Class Analogy
SAC	Strong Alkalized Cocoa
SD	Standard Deviation
SEC	Standard Error of calibration

Abbreviations and Acronyms

Si-BPANNR	Synergy Interval Backpropagation Neural Networks Regression
Si-GAPLS	Synergy Interval-Genetic Algorithm-PLS
SNV	Standard Normal Variate
SPME	Solid-Phase Micro-Extraction
SVMR	Support Vertor Machine Regression
TE	Trolox equivalents
THz	Terahertz
TFC	Total Fungi Count
TMP	Tetramethylpyrazines
TOFMS	Time of Fly Mass Spectrometry
TPC	Total Polyphenol content
TRAPP	Total Radical Trapping
TrMP	Trimethylpyrazines
UHPLC	Ultra-High-Performance-Liquid Chromatography
UFC	Unfermented cocoa
UPLC-MS/MS	Ultra high performance liquid chromatography-tandem mass spectrometer
UV	Ultraviolet
VCEAC	Vitamin C equivalence antioxidant capacity
VFA	Volatile Fatty Acids
VIP	Variable Importance in Projection
XRD	X-ray diffraction
2nd Der	Second derivative

INDEX

1. INTRODUCTION

Introduction is based on the article: Quelal-Vásconez, M. A., Lerma-García, M.J., Pérez-Esteve, É., Talens, P., & Barat, J. M. (2020). Roadmap of cocoa quality and authenticity control in the industry: a review of conventional and alternative methods. Submitted to Comprehensive Reviews in Food Science and Food Safety.

Roadmap of cocoa quality and authenticity control in the industry: a review of conventional and alternative methods

Maribel Alexandra Quelal Vásconez,[a] María Jesús Lerma-García,[a] Édgar Pérez-Esteve,[a] Pau Talens,[a] José Manuel Barat,[a]

[a]Departamento de Tecnología de Alimentos, Universitat Politècnica de València. Camino de Vera s/n, 46022, Spain

1

Abstract

Cocoa (*Theobroma cacao* L.) and its derivatives are appreciated for their aroma, color and healthy properties, and are commodities of high economic value worldwide. Wide ranges of conventional methods have been used for years to guarantee cocoa quality. Recently however, demand for global cocoa and the requirements of sensory, functional and safety cocoa attributes have changed. On the one hand, society and health authorities are increasingly demanding new more accurate quality control tests, including not only the analysis of physico-chemical and sensory parameters, but also determinations of functional compounds and contaminants (some of which come in trace quantities). On the other hand, increased production forces industries to seek quality control techniques based on fast, nondestructive online methods. Finally, an increase in global cocoa demand and a consequent rise in prices can lead to future cases of fraud. For this reason, new analytes, technologies and ways to analyze data are being researched, developed and implemented into research or quality laboratories to control cocoa quality and authenticity. The main advances made in destructive techniques focus on developing new and more sensitive methods chromatographic analysis to detect metabolites and contaminants in trace quantities. These methods are used to: assess cocoa quality; study new functional properties; control cocoa authenticity; or detect frequent emerging frauds. Regarding to non destructive methods, spectroscopy is the most explored nondestructive technique, which is conducted within the near infrared range, and also within the medium infrared range to a lesser extent. It is applied mainly in the postharvest stage of cocoa beans to analyze different biochemical parameters or to assess the authenticity of cocoa and its derivatives.

Keywords: Cocoa quality roadmap, Chemometrics, authenticity control, nondestructive methods, multivariate analysis.

Introduction

Practical Application:

Recent trends in the cocoa sector (increased quantity and quality demands, new technical specifications, emerging functional properties, global food quality control trends, such as fast, non-destructive online methods) mean that the cocoa industry has new analysis requirements. This work aims to guide researchers and quality control technicians to the possibilities available today to control cocoa quality and authenticity in the fastest most reliable way to make cocoa production more efficient, safe, fast and innovative.

4

1. Introduction

Cocoa (*Theobroma cacao* L.) is a commodity of high economic value worldwide. Most of its production comes from West African countries (mainly the Ivory Coast and Ghana, which account for approximately 60% of the world's total cocoa), but is usually processed in the European Union (1.3 million tons or 40% of the global processing market in 2014). Apart from the Europen Union, cocoa beans are also processed in Indonesia, EEUU and Brazil in significant quantities (CBI, 2016; Shavez, Ahmad, Jan, & Bashir, 2017; ICCO, 2019).

In the different producing areas, three main distinct varieties are produced. The most ancient and most appreciated chocolate manufacturer variety is called Criollo (which means native), and is that traditionally cultivated by the Aztecs and Mayans in Central and South America. Later a new variety that better resists diseases and pests, called Forastero (meaning foreign), was taken from Amazon regions to other cocoa-growing areas in Latin America, and was exported to other West Africa and East Asia countries. Finally, in order to combine the advantages of Forastero and the appreciated fine flavor of Criollo, a new hybrid variety was harvested, known as Trinitario. Besides these varieties, the Nacional variety, which is generally considered native to Ecuador, is receiving more attention in the cocoa market for its sensory properties (Crouzillat *et al*., 2000). Each variety has specific sensorial characteristics that are related to its origin, environmental conditions and fermentation (Chetschik *et al*., 2018; Loullis & Pinakoulaki, 2018). Forastero is considered a bulk variety, while Criollo, Trinitario and Nacional are considered fine varieties. Bulk cocoas usually possess strong harsh flavors, while fine cocoas are perceived as being more aromatic or smoother (Counet *et al*., 2004). Growing conditions and postharvest practices can condition the final features of cocoa pods and, thus, of cocoa products (ADM Cocoa Manual, 2006). Therefore, knowing the variety and geographical indication of the cocoa beans used as raw material to produce different cocoa products is becoming increasingly more important as it can condition the final quality and, hence, cocoa prices.

5

Introduction

Regardless of cocoa variety, cocoa beans are subjected to different postharvest and industrial processes to obtain distinct cocoa products (Di Mattia *et al*., 2014, Aprotosoaie, Luca, & Miron, 2016). The first steps include cocoa bean fermentation and drying (Suazo, Davidov-Pardo, & Arozarena, 2014). Next fermented and dried cocoa beans undergo several industrial processes. Bean shelling provides nibs and the first subproduct: shells (Tan & Kerr, 2018). Nibs can be roasted and milled to obtain cocoa liquor (Ioannone *et al*., 2015). When cocoa liquor is pressed, two products are obtained: cocoa butter and cocoa cake (Oliviero, Capuano, Ca, & Fogliano, 2009). Finally, cocoa cake undergoes another milling step to provide cocoa powder. Optionally, another important step to develop color and flavor, called alkalization or dutching process, can be performed in different cocoa products: cocoa nibs, cocoa cake or cocoa powder (Pérez-Esteve, Lerma-García, Fuentes, Palomares, & Barat, 2016). Alkalization is normally carried out by adding sodium or potassium carbonate at high temperature and controlled pressure. According to the final pH, cocoa powders can be classified into natural (pH 5-6), light-alkalized (pH 6-7.2), medium-alkalized (pH 7.2-7.6) and strong-alkalized powders (pH > 7.6) (Miller *et al*., 2008). Light-alkalized cocoa powders are light brown, but darker than natural ones, and their flavor is less astringent and less acidic than those of natural powders. Strong-alkalized cocoa powders are very dark and have a much stronger flavor than medium-alkalized ones (Kostic, 1997). A summary of all these processes is shown in Figure 1.

If cocoa bean quality is poor, the quality of the final products will be worse. So over the years, the cocoa industry has defined different relevant aspects, such as the physical characteristics with a direct bearing on manufacturing performance or flavor which, over time, have become the commercial standards employed worldwide.
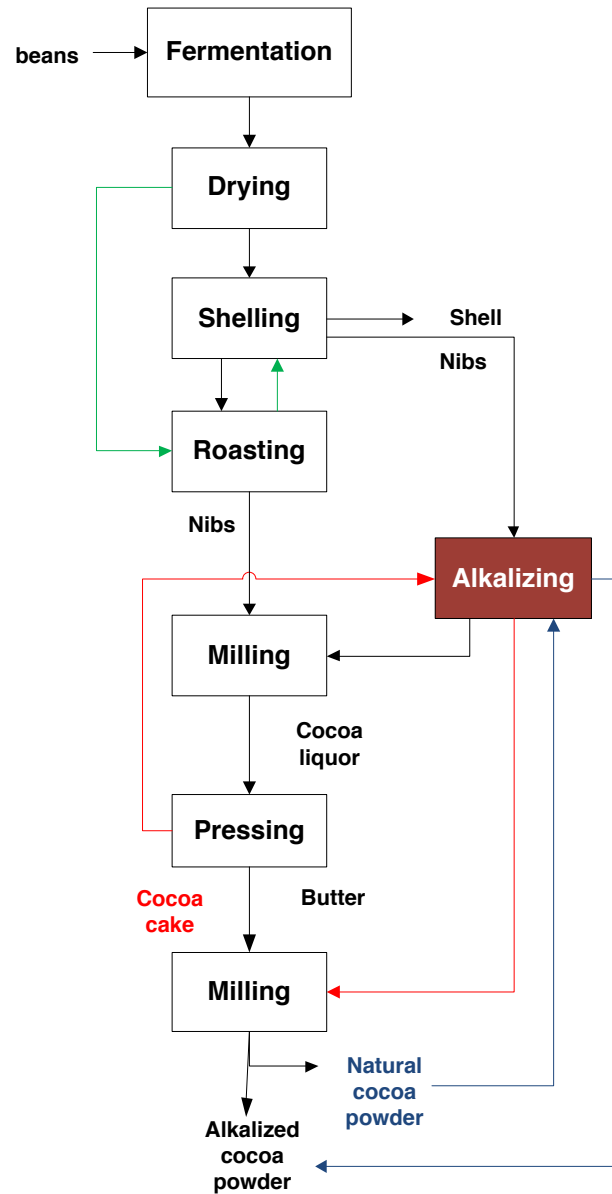
Figure 1. Cocoa and derivatives flow processing chart. Alkalization ways: black (nibs), red (cocoa cake), natural cocoa powder (blue).

These commercial standards for cocoa beans, cake or chocolate usually include parameters related to physico-chemical parameters and compositional features (see Table 1). These evaluations aim to obtain a product that combines ideal aroma, flavor, color, technological behavior and functional compounds.

Introduction

Table 1: General quality attributes of cocoa beans, chocolate and cocoa powders.

Quality attributes	Details	Observation
Cocoa beans		
Size (#beans/100g)	≤ 100	Standard beans
	101-110	Medium beans
	111-120	Small beans
	> 120	Very small beans
Uniformity	Variable-sized beans are harder to break and deshell	
Fermentation	5% slaty, 5% defectiveness	Good fermented
	10% slaty, 10% defectiveness	Fair fermented
Moisture	<8%	Aceptable
Chemical residues	According to authority regulations	Under limits
Chocolate and cocoa powders		
Fat content		Characteristic
Fat quality	Low in free fatty acids, show characteristic melting and solidification properties	
Aroma and flavor	Without moldy off-flavors, smoky taints, acidic off-flavors, proximity to another strong-smelling products	Characteristic
Color		Characteristic
Cocoa powder		
Solubility	95%	Good solubility
Shell content	< 5% in fat free-dry cocoa	Aceptable

CAOBISCO-ECA-FCC (2015), ADM Cocoa Manual (2006)

     This goal is fulfilled by assessing the physico-chemical cocoa characteristics in raw material and its derivatives in each processing stage (Miller *et al.*, 2006). Indeed, each processing stage comprises key quality control processes that should be addressed to obtain high quality cocoa products. For example, the fermentation

8

control in the postharvest stage is crucial for the formation of aromatic compounds to form (Aculey *et al.*, 2010), and further quality control points should be set to guarantee quality requirements (e.g. fat content, moisture, etc.) while drying, industrial roasting and alkalization cocoa processes.

Apart from its nutrients, technological properties pleasant flavor, aroma and color, cocoa is also known for offering many health benefits (Bonvehí, 2005) because it is an excellent source of antioxidants (Langer, Marshall, Day, & Morgan, 2011). Many different bioactive compounds are present in cocoa, such as polyphenols, mainly flavonoids (flavanols, procyanidins, and anthocyanins) and methylxanthines (caffeine, theobromine) (Talbot, Mensink, Smolders, Bakeroot, & Plat, 2018), among others. These phytochemicals can be present at different concentrations depending on diverse factors like cocoa variety and cocoa processing, which can lead to the presence of new bioactive compounds. For example, cocoa roasting is a precursor for the formation of heterogeneous high-molecular-weight polymers known as "melanoidins", which are related to antihypertensive and antioxidant properties (Quiroz-Reyes & Fogliano, 2018).

Cocoa phytochemicals are an excellent ally to prevent cardiovascular and other chronic diseases, which are the main cause of mortality in Western countries (Gianfredi, Salvatori, Nucci, Villarini, & Moretti, 2018; Martín & Ramos, 2017). It has been shown that cocoa's lipid profile balance is beneficial given the presence of stearic acid, which is a saturated fatty acid present in high proportions in cocoa butter (ca. 35%). The behavior of this fatty acid is unusual because, despite being a saturated fat, it behaves like an unsaturated one and has a neutral effect on blood cholesterol levels (Torres-Moreno, Torrescasana, Salas-Salvadó, & Blanch, 2015).

Polyphenols, especially epicatechin, perform neuroprotective and neuromodulatory action. The former action is associated with the prevention and reduction of neurological, cognitive and functional brain diseases (Alzheimer's, Parkinson's and senile dementia). The second action is related to cognition, humor, learning and memory skills (Ishaq & Jafri, 2017). These healthy cocoa benefits

promote its employment as a basic ingredient used by the pharmaceutical and cosmetic industries (APEDA, 2015; Oracz, Nebesny, & Żyżelewicz, 2015).

Based on cocoa's, and therefore on chocolate's. sensory attributes and functional properties, it is not surprising that global cocoa demand is increasing. Demand for cocoa is predicted to rise by 30% by 2020, which equals the present production output of the Ivory Coast (1 million tons) (Afoakwa, Quao, Takrama, Budu, & Saalia, 2013; Shavez *et al*., 2017). The extent of this growth is such that without empowering and investing in small-scale farmers, the industry will struggle to provide sufficient supply. This increasing cocoa demand, volatile prices and the uncertain global cocoa production, which is at risk due to climate change, can lead to cases of cocoa adulteration.

In this context, the development of new and faster analysis methods is not only essential for guaranteeing quality specifications and costumers requirements, or for process control purposes; but also important to explore new properties of cocoa products and to detect new frauds attempting food safety and cocoa authenticity. Therefore, the goal of this review is to provide a comprehensive insight into both traditional and fast nondestructive technologies that might be used in the cocoa industry to assess cocoa composition and quality, to study new cocoa properties and to detect frequent and emerging frauds.

2. Determination of cocoa components

2.1 Major components

Cocoa compounds, such as fat, nitrogenous compounds, protein, moisture, ash and fiber, are usually evaluated by proximate analyses. Fat is determined by the AOAC 963:15 Method, which consists in a Soxhlet extraction method, moisture is determined by the AOAC 931:04 method, protein by measuring the nitrogen content with the Kjeldahl method (AOAC 970:22), ash by the AOAC 972:15 method and

fiber by AOAC 991.43. An example of a recent application of these methods is a study about the effect that solar heat has on cocoa beans (Abdullahi, Muhamad, Dzolkhifli, & Sinniah, 2018). Automation improvements of these methods have been incorporated into both industry and R&D laboratories. For instance, fat can be determined by the Soxtec™ AVANTI 2050 system (Servent *et al*., 2018; Sess-Tchotch *et al*., 2018), while nitrogen content can be determined using an automatic Kjeldahl apparatus (Hue *et al*., 2016) or a micro-Kjeldahl apparatus, which allow microquantities to be established (Hashimoto *et al*., 2018). With these analyses, it is possible to establish the nutritional information of cocoa and derivatives, which is usually included on product labels.

By applying these techniques, it can be concluded that fat, nitrogenous compounds, sugars and polyphenols are the main constituents of cocoa products. Cocoa fat is roughly 57%, 6.6%, and 11%, and total nitrogen content is ca. 2.5%, 3.2% and 4.3% for nibs, cocoa shells and cocoa powder, respectively. The percentage of water is ca. 3.2% in nibs, 6.6% in cocoa shells and 3% in cocoa powder (Afoakwa *et al*., 2013; ICCO, 2012). Cocoa powder also contains a relevant polysaccharide content (comprising cellulose, hemicellulose, and pectin), noncarbohydrate lignin, nonstructural polysaccharides like gums and mucilage. It also contains considerable amount of flavanols and organic acids (ca. 4% among lactic and acetic acids), which are responsible for cocoa color (Shavez *et al*., 2017). Table 2 summarizes the main components of cocoa powders.

Table 2: Cocoa powder composition (ICCO, 2012; Krähmer *et al*., 2015; Lacueva *et al*., 2008).

Component	Major compounds (%)
Fat	11
Moisture	3
Total nitrogen	4.3
Nitrogen (corrected for alkaloids)	3.4
Protein	20
Nitrogen corrected for alkaloids x 6.25 %	21.2
Ash	5.5
Water soluble ash	2.2
Phosphate (as P_2O_5)	1.9
Ash insoluble in 50% HCl	0.08

	Minor compounds (mg/g)
Flavanols	
Catechin	0.6
Epicatechin	5.7
Methylxanthines	
Caffeine	6
Theobromine	28
Other compounds:	
Total procyanidins	22
Total amino acids	3.4
Total sugars	8.9

*Reference mean values, since both flavanols and methylxanthines contents are highly influenced by cocoa origin, postharvest and processing processes

2.2 Bioactive compounds: Polyphenols and methylxanthines

Polyphenols are the most relevant bioactive cocoa compounds found to date. They can be divided into at least 10 different classes depending on their basic

structure in the plant kingdom (Wollgast & Anklam, 2000). In cocoa, the flavonoids family constitutes the most important single group, which can be further divided into several classes. The main classes of flavonoids found in cocoa are summarized in Figure 2.
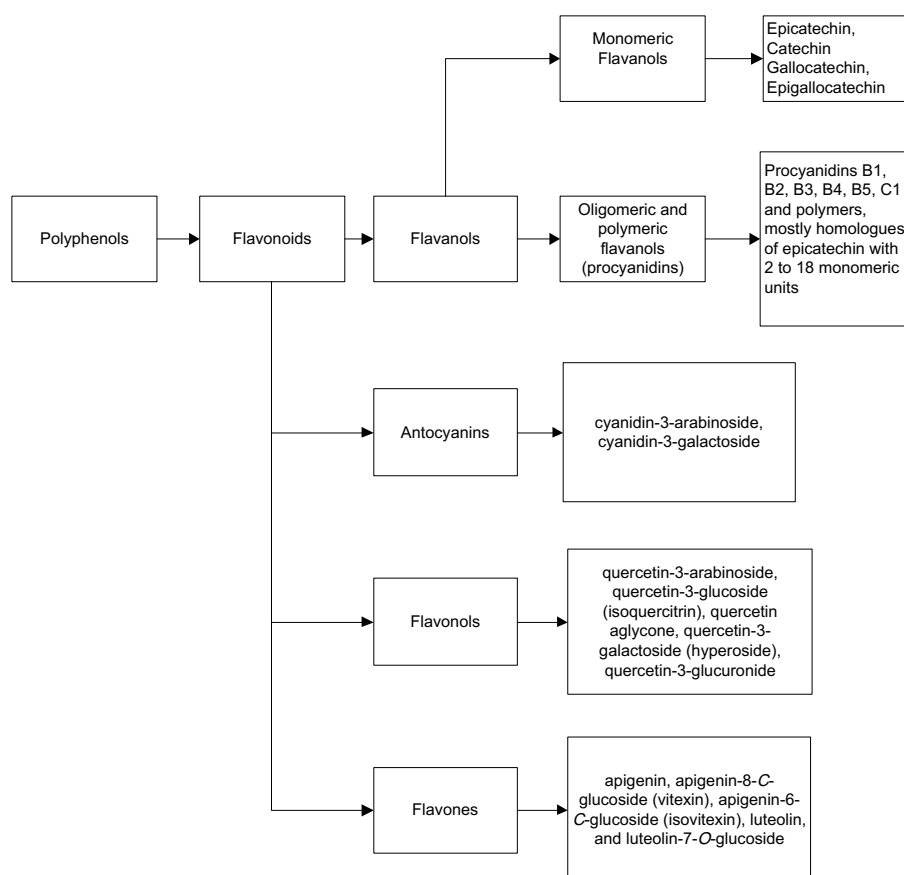
Figure 2: Summary of the main polyphenol classes found in cocoa.

The most abundant flavonoids in cocoa comprise monomeric flavanols (or flavan-3-ols) and their oligomeric and polymeric forms (known as procyanidins) (Lacueva *et al.,* 2008). Monomeric flavanols include epicatechin (reported as the major monomeric flavanol in cocoa that represents ca. 35% of total phenolic content (Lacueva *et al*., 2008), catechin (found in smaller amounts) and also traces of gallocatechin and epigallocatechin (Wollgast & Anklam, 2000). Procyanidins, also known as condensed tannins, are mostly flavan-3,4-diols, which are $4 \rightarrow 8$ or $4 \rightarrow 6$ bound to condensed dimers, trimers or oligomers with epicatechin as the main

extension subunit (Wollgast & Anklam, 2000). In cocoa, procyanidins with a degree of polymerization (DP) up to decamer have been identified. Oligomers (procyanidins B1, B2, B5, and C1) and polymers account for 90% of total polyphenols, while monomers account for 5–10% (Lacueva *et al*., 2008). Another flavonoid class is anthocyanins, which is the most important group of water-soluble plant pigments responsible for the color of flowers and fruits of higher plants (Wollgast & Anklam, 2000). The main anthocyanins identified in cocoa beans are cyanidin-3-arabinoside and cyanidin-3-galactoside, which represent ca. 4% of the total polyphenol content of cocoa beans, but they can be hydrolyzed during the cocoa fermentation process (Forsyth & Quesnel, 1957; Wollgast & Anklam, 2000).

Other important bioactive compounds found in cocoa and cocoa products are methylxanthines (Li *et al*., 2012). The main methylxanthines present in cocoa include caffeine and theobromine, but low levels of theophylline have also been found. These compounds are related to psychoactive properties that lead to better daily human life (i.e., more efficient thinking, exploring, hunting, etc.) without the serious side effects of drugs of abuse (Franco, Oñatibia-Astibia, & Martínez-Pinilla, 2013).

Both polyphenol and methylxanthine compounds are responsible for the astringent and bitter taste of cocoa, which affects cocoa stability and digestibility (Li *et al.,* 2012). Moreover, they are generally determined to control the quality of the cocoa products obtained from raw beans in all the processing steps until end (ready-to-eat) products are obtained. Therefore, their determination is very important for the cocoa industry.

Phenolic compounds are usually extracted from cocoa matrices using different solvents, and methanol is considered the most efficient one (Belščak, Komes, Horžić, Ganić, & Karlović, 2009), although other solvents solutions, like acetone, water and acetic acid, are also widely used. The polyphenol content of cocoa is usually evaluated by total polyphenol content (TPC), antioxidant capacity (which can be

obtained by different assays, which are described below), and by also quantifying the different individual polyphenols present in samples.

TPC is usually determined by the Folin-Ciocalteu colorimetric assay, which is based on the Folin-Ciocalteu reagent's ability to react with phenolic hydroxyl groups (Manzano *et al*., 2017).

Antioxidant capacity can be established by methods based on both hydrogen atom or electron transfer reactions. The first category includes methods like ORAC (oxygen radical absorbance capacity), TRAP (total radical trapping antioxidant parameter), Crocin bleaching assay, IOU (inhibited oxygen uptake), inhibition of linoleic acid oxidation and inhibition of LDL (Low Density Lipoprotein) oxidation. The second category includes assays such as TEAC (Trolox equivalent antioxidant capacity), FRAP (ferric ion-reducing antioxidant parameter) DPPH (diphenyl-1-picrylhydrazyl), copper (II) reduction capacity, etc. (Di Mattia *et al*., 2014). The heterogeneous methods (different reagents) used to determine antioxidant activity make the comparison of the obtained results difficult. However, the most frequently used assays are ABTS, DPPH, ORAC, TRAP, and FRAP. These methods can provide discordant results depending on the most abundant antioxidant molecules in the system and their interactions (Di Mattia *et al*., 2014).Individual determinations of both polyphenols and methylxanthines are usually performed by HPLC-UV, but the concomitant identification of other unknown compounds, mainly flavan-3-ol derivatives (Fayeulle *et al*., 2018), has led to the proliferation of innovative, miniaturized and/or two-dimensional HPLC methodologies (Toro-Uribe, Montero, López-Giraldo, Ibáñez, & Herrero, 2018). For this purpose, other detectors like mass spectrometry are widely used (Cádiz-Gurrea *et al*., 2014; Pedan *et al*., 2016; Rodríguez-Carrasco, Gaspari, Graziani, Sandini, & Ritieni, 2018).

Many articles have been published in the literature about the determination of and/or the changes produced in the different types of polyphenols and methylxanthines among several distinct cocoa products (Gabbay Alves *et al*., 2017; Machonis, Jones, Schaneberg, Kwik-Uribe, & Dowell, 2014; Manzano *et al*., 2017;

Introduction

Risner, 2008), in cocoa processing steps (Elwers, Zambrano, Rohsius, & Lieberei, 2009; Lacueva *et al.*, 2008; Li *et al.*, 2012, 2014; Miller *et al.*, 2008; Payne, Hurst, Miller, Rank, & Stuart, 2010; Pedan, Fischer, Bernath, Hühn, & Rohn, 2017; Quiroz-Reyes & Fogliano, 2018), between different cocoa clones or varieties (Elwers *et al.*, 2009; Niemenak, Rohsius, Elwers, Omokolo Ndoumou, & Lieberei, 2006), etc. Therefore, some of these studies are reviewed below.

Risner (2008) determined both methylxanthines (theobromine and caffeine) and flavan-3-ols (catechin and epicatechin) by HPLC in different cocoa products, including standard reference material baking chocolate 2384, cocoa powder, cocoa beans, and cocoa butter.

Miller *et al.* (2006) published a study in which antioxidant capacity (the ORAC method), vitamin C equivalence antioxidant capacity (VCEAC), TPC and procyanidin contents were determined and analyzed by principal component analyses (PCA) to identify their behavior in different cocoa derivatives, such as natural cocoa powders, unsweetened baking chocolates, semisweet baking chips, milk and dark chocolates and chocolate syrups. The highest levels of antioxidant activities, TPC and procyanidins were found in natural cocoa powders, followed by baking chocolates, dark chocolates, baking chips, and by finally milk chocolate and syrups.

In another study, the influence of alkalization on TPC, methylxanthines, flavan-3-ols and other components, such as volatiles, free amino acids, and sugars, was studied in commercial cocoa powders (Li *et al.*, 2012). The results showed that the content of both methylxanthines and flavan-3-ols lowered as the degree of alkalization increased, while a higher degree of alkalization decreased TPC. Similar results were found by Miller *et al.* (2008), who also studied the influence of alkalization on the antioxidant capacity (ORAC method), TPC and flavanol content of cocoa powders. For all the samples, the highest contents of all the determinations were found for natural powders.

16

The influence of the alkalization process on the content of both monomeric flavanols (catechin and epicatechin) and flavonols (quercetin-3-glucuronide, quercetin-3-glucoside, quercetin-3-arabinoside, and quercetin) in cocoa powders was studied by Lacueva *et al.* (2008). The authors concluded that the marked reduction found in the flavonoid content of natural cocoa powder, together with the change observed in the monomeric flavanol profile that resulted from alkalization treatment, could affect the antioxidant properties and the polyphenol bioavailability of cocoa powder products.

Li *et al.* (2014) studied the effects of alkalization treatments on color, colorimetric fractions, TPC, and anthocyanin contents of cocoa powders. They concluded that the color qualities of cocoa powder can be improved by optimizing alkalization parameters. For example, cocoa powders alkalized with K_2CO_3 displayed darker colors and lower TPC than the powders alkalized with NaOH. High temperature and basic pH conditions gave a darker color due to sugar degradation, Maillard reactions and anthocyanin polymerizing.

In addition to the changes that take place during alkalization, further studies have studied the influence of other processing steps. One such case is the work published by Quiroz-Reyes *et al.* (2018), who evaluated the effect that roasting and fermentation steps had on TPC, and antioxidant capacity and proanthocyanidins, melanoidins and flavan-3-ols contents on two cocoa bean varieties (Forastero and Criollo). The results showed that the Forastero variety was characterized by the highest melanoidins content, antioxidant capacity (DPPH Quencher assay) and TPC values under severe roasting conditions, while severer thermal treatments lowered the concentration of TPC and proanthocyanidins in both varieties, and also influenced the flavan-3-ols profile. Thus it can be concluded that a proper roasting process design and adequate cocoa variety selection can optimize the cocoa health potential, especially melanoidins and phenolic compounds.

In another study (Payne *et al.*, 2010), the impacts of fermentation, drying, roasting and alkalization processes on catechin and epicatechin contents were

evaluated in both unfermented and fermented cocoa beans. The results showed that unripe cocoa beans had a 29% higher level of epicatechin and the same level of catechin as fully ripe beans, while no significant difference in the content of both flavanols was observed during drying. A marked reduction (>80%) in catechin and epicatechin levels was observed in fermented *versus* unfermented beans. During roasting, loss of epicatechin took place along with a concomitant increase in the catechin level, probably due to the epimerization of epicatechin. Finally, alkalization led to a reduction in both catechin and epicatechin contents. Therefore, these authors proposed using the epicatechin/catechin ratio as a useful sensitive indicator for the processing history of cocoa beans.

Pedan *et al*. (2017) studied the influence of different lab-scale chocolate manufacturing process stages (including opening fresh cocoa pods, fermentation, drying, roasting and conching, and finishing chocolate bars) on the content of oligomeric proanthocyanidins and their antioxidant capacity by the NP-HPLC-online-DPPH methodology. For this purpose, one single batch of 5 kg of fresh Trinitario variety cocoa beans was studied in the different processing stages. The results showed that the total proanthocyanidin content continuously lowered during the manufacturing process, with only ca. 20% of the initial content present in chocolate.

As previously indicated, several studies have been conducted in which the influence of cocoa clones, variety and/or origin on polyphenols content has been studied (Elwers *et al.,* 2009; Niemenak *et al.*, 2006). For example, Niemenak *et al*. (2006) compared TPC, flavanol (catechin and epicatechin) and anthocyanin (cyanidin-3-galactoside and cyanidin-3-arabinoside) contents of different seeds from Cameroon. The obtained results suggested that there was no qualitative difference in TPC in cocoa beans despite their genetic origin and fermentation-like process. However, a quantitative difference in epicatechin, catechin, cyanidin-3-galactoside and cyanidin-3-arabinoside, and also in three undefined substances, was found. This difference was attributed to growing conditions (microclimate, position of pods on

trees, etc.). Finally, PCA and hierarchical cluster analyses classified samples according to their polyphenol and anthocyanin contents.

Alternative methods for analyzing these bioactive compounds (polyphenols and methylxanthines) are included in Section 4.

2.3 Fatty acids

It has been reported that cocoa beans and cocoa liquor have around 50g/100 g of fat (Hashimoto *et al*., 2018). This fat, also called cocoa butter, is frequently reported to be the main vegetable fat used in chocolate manufacturing due to its rheological, textural and chemical characteristics, such as triglycerides and fatty acids (FA) composition (Guehi *et al*., 2008). Cocoa butter hardness depends on the ratio between saturated and unsaturated fatty acid bound in triglycerides, and on the free fatty acids (FFA) content. Whereas cocoa butter hardness increases with a higher proportion of saturated fatty acids, higher FFA content reduces this parameter. Thus Council Directive 73/241/EEC (EU, 2000) limits maximum FFA contents to a 1.75% oleic acid equivalent in cocoa butter (Guehi *et al*., 2008). The FA profile is also linked to cocoa aroma quality as the presence of volatile fatty acids (e.g. acetic, propionic, butric, isobutric, and iso-valeric acids) is linked to low quality products (García-Alamilla *et al*., 2007). Then there is stearic fatty acid (C18:0), which offers health benefits (Torres-Moreno *et al*., 2015). In this context, the characterization of both the quantity and quality of FA present in cocoa seeds and cocoa products is important and frequently evaluated (Guehi *et al*., 2008).

In cocoa butter, total FFAs are determined by measuring the amount of base needed to neutralize oleic acid (titration method) according to the official method 42-1993 (IOCCC, 1996). This method consists in dissolving 5 g of extracted cocoa butter in 50 ml of a previously hot petroleum ether/absolute ethanol mixture (1:1, v/v) neutralized by adding phenolphthalein. The mixture is then titrated with 0.1N alcoholic KOH solution. This method was used by Guehi *et al*., (2008) to study how

storage conditions affect the FFA contents of raw cocoa beans. The above-cited authors used different samples of fermented-dried cocoa beans purchased from the Ivory Coast. The authors reported very low FFA contents (0.2-0.8%) in whole healthy cocoa. Their study also stated that FFA formation did not depend on either genotype or cocoa post-harvest processing technologies (number of fermentation days). However, in defective cocoa beans, high and increasing FFA contents were found. This increased content was attributed to the activity of microflora, which has been associated with initial quality and loss of the physical integrity of cocoa beans.

The FA profile can be determined by preparing FA methyl esters (FAMEs) using method AOAC 948.22 and gas chromatography coupled to mass spectrometer detector GC-MS (Torres-Moreno *et al*., 2015). By the aforementioned method, Torres-Moreno *et al*., (2015) studied the influence of the geographical origin (Ecuador and Ghana) and processing conditions of chocolate (three roasting times: 30.5, 34.5 and 38.5 min; two conching times: 24 and 42 h) on the FA profile. For this purpose, the authors used the official method 948.22 (AOAC International, 1990b) and identified 15 FA in cocoa and chocolates. Of these, the most important FA were C16:0 (>25%), C18:0 (>33%) and C18:1 (>32%), expressed as the relative percentage of the total fatty acid content in unroasted cocoa beans and in the chocolate made from Ecuadorian and Ghanaian samples. For cocoa, differences in the FA profile were found in C12:0, C14:0, C16:0, C16:1, C17:0, C17:1 and C18:0, while differences were found only in C16:0, C18:0, C18:1 and C18:2 for chocolates. For all the samples, C16:0, C18:0, C18:1 and C18:2 were quantitatively the most important FA. Differences in the FA profile were explained mainly as an effect of the geographical origin and were not due to processing conditions in chocolate. Thus Ecuadorian chocolate showed a healthier FA profile with larger amounts of unsaturated FA and smaller amounts of saturated FA than Ghanaian chocolate.

2.4 Amino Acids

Amino acids take part in the aroma and flavor formation of cocoa and cocoa-related derivatives (Voigt, Textoris-Taube, & Wöstemeyer, 2018). Their content is also related to human health (Stark, Lang, Keller, Hensel, & Hofmann, 2008). Thus, in addition to total protein contents, knowing the profile of the amino acids that form these proteins is essential.

High-performance liquid chromatography is the method normally used to analyze amino acids. As amino acids do not exhibit chromophore groups in their structure, they cannot be detected by UV–VIS spectrometry. Thus they have been traditionally derivatized before being analyzed. During the derivatization step, a UV–VIS nonresponding analyte can be converted into a compound with significant absorbance or fluorescence that allows determinations with greater sensitivity (Kubíčková *et al.*, 2011).

One study that aimed to correlate amino acid content with cocoa aroma was published by Voigt *et al.*, (2016). These authors analyzed amino acid content in cocoa beans to characterize the amino acid sequence of aroma precursor peptides. For this purpose, amino acids were converted into their *o*-phthalaldehyde (OPA) derivatives and then separated by reversed-phase HPLC. Effluents were monitored fluorometrically. Another study using derivatization with a fluorescent chromophore to quantify the content of free amino acids in Forastero cocoa beans was conducted by Hinneh *et al.*, 2018. In this work, the authors evaluated the influence of pod storage on the free amino acid profiles and the implications on the development of some Maillard reaction related to flavor volatiles. As a result, they found that although the concentration of free amino acids was directly proportional to pod storage duration, significant differences were observed for pod storage periods exceeding 7 days (Hinneh *et al.*, 2018).

In relation to health properties, amino acids and their metabolites can act as functional molecules. Kynurenic acid, obtained during the metabolization of amino

acids like tryptophan through the kynurenine pathway, exhibit antioxidant capacity. Several authors have attempted to quantify tryptophan content and its derivatives in the kynurenine pathway by liquid chromatography with various detectors. One study that analyzed tryptophan and its derivatives in the kynurenine pathway in cocoa is that reported by Yılmaz and Gökmen, 2018. In their study, the authors compared the content of these analytes in several fermented food products (bread, beer, red wine, white cheese, yogurt, kefir and cocoa powder). Tryptophan derivatives were determined by ultra-high-performance liquid chromatography-tandem mass spectrometer (UPLC–MS/MS). Of these analytes, cocoa powder contained more kynurenic acid, which is a neuroprotective compound (Yılmaz & Gökmen, 2018).

The aim of another recent application of cocoa amino acids quantification was to assess the geographical origin (Asia, Africa and South America) of cocoa beans used to produce chocolate (Acierno, Alewijn, Zomer, & van Ruth, 2018). For this purpose, the authors tested the applicability of Flow Infusion-Electrospray Ionization-Mass Spectrometry (FI-ESI-MS). Among the tentatively identified compounds, the authors recognized free amino acids that could be used to distinguish the geographical origin of cocoa beans. This fell in line with other studies that have reported the geographical influence on the free amino acid concentration in raw cocoa (Rohsius, Matissek, & Lieberei, 2006).

2.5 Peptides

As with amino acids, the presence and concentration of certain peptides (e.g. N-terminal 15-kDa vicilin found in South American CCN51 samples) can be used to evaluate the origin of a particular cocoa. Kumari *et al*., (2018) used ultra-high-performance liquid chromatography-electrospray ionization mass spectrometry (UHPLC-ESI-Q-q-TOF-MS) to analyze the proteins and oligopeptides of nonfermented and fermented beans of various geographic origins. ESI is a soft ionization method capable of providing both protonated and deprotonated molecules.

Q-TOF-MS is able to combine high sensitivity and mass accuracy for both precursor and product ions and, therefore, allows the elemental composition for both parent and fragment ions to be confirmed both quickly and efficiently. UHPLC can provide high resolutions for the separation of complicated natural products and improves the sensitivity of Q-TOF-MS detectors (Li *et al.*, 2017). In this study, the authors observed how protein quantities, and their profiles that derived from two-dimensional gel electrophoresis, showed striking differences for nonfermented beans depending on their geographical origin. However, in fermented beans, the detected diversity of peptides did not correlate with geographical origin, but to the degree of fermentation. These findings suggest that the variability in peptide patterns depends on the fermentation method applied in the country of origin, which ultimately indicated diversified proteolytic activities (Kumari *et al.*, 2018).

2.6 Sugars

Cocoa sugars are cocoa aroma precursors that are present in higher proportions in cocoa pulp as fermentable sugars (9-13% w/w). The predominant sugars in cocoa beans are sucrose, fructose and glucose (Afoakwa, 2010). In cocoa beans, fermentation allows reducing sugar (fructose and glucose) formation. Therefore, during the roasting process they undergo Maillard reactions and Strecker degradation, which lead to the generation of desirable flavor volatiles. Thus reducing sugars determination is important for cocoa sensorial control purposes (Kongor *et al.*, 2016).

A traditional method to analyze total and reducing sugars in cocoa beans and products is that known as the phenol sulfuric acid method (Dubois, Gilles, Hamilton, Rebers, & Smith, 1956). This method allows simple sugars, oligosaccharides, polysaccharides, and their derivatives, to be detected, including methyl ethers with free or potentially free reducing groups as they give an orange-yellow color after treatment with phenol and concentrated sulfuric acid.

However, the identification and quantification of different reducing sugars require a more selective technique. One common alternative is to use gas chromatography after aqueous extraction and derivatization. Hinneh *et al.* (2018) analyzed the sugar profile of Forastero cocoa beans by gas chromatography. For this purpose, these authors obtained an extract that was then derivatized in two steps: first oximation and second the formation of trimethylsilylesters. The study revealed that on storage day 0, cocoa pods exhibited 0.672±0.004 g/100 g of fructose, 0.264±0.001 g/100 of glucose and 0.021±0.001 g/100g of sucrose. These amounts varied with storage. After 3 pod storage days, the amount of glucose and sucrose had increased. After 7 pod storage days, these amounts lowered, while the amount of fructose increased, so the respective fructose-glucose ratios for 0 PS, 3 PS, and 7 PS were approximately 3:1, 2:1 and 4:1. This confirms the role of PS in influencing sugar degradation dynamics through nib acidification during fermentation.

2.7 Aroma and flavor

Aroma and flavor are the most appreciated cocoa bean features as they contribute to the final flavor of chocolates and other derived products. Samples can be evaluated for cocoa strength or chocolate flavor, residual acidity, bitterness and astringency, and for the presence of any off-flavor and positive ancillary flavors, such as fruity or floral. The sensory evaluation of cocoa products can be made by difference and descriptive tests. Difference tests are performed to compare samples, or samples against a standard, which include the triangle test, paired comparisons, ranking and the two-out of five test. No expert training is needed to carry out these tests (ADM Cocoa manual, 2006). Descriptive tests include the flavor profile method (FPM), the descriptive analysis test (QDA) and the free choice profiling (FCP), a variant of (QDA). Sensorial analysis methods may also include the use of a principal component analysis (PCA), which allows variable reduction according to inter-related connections. The information displayed in a two-dimensional graph provides

essential information on the flavor profiles of cocoa samples based on descriptors. This method was used by Luna *et al*., (2002) to evaluate the flavor of Ecuadorian cocoa liquor, who concluded that polyphenols could be essential for the overall perception of cocoa liquor characteristics (CAOBISCO-ECA-FCC, 2015; Luna, Crouzillat, Cirou, & Bucheli, 2002).

Aroma and flavor are conditioned by different parameters that are chemical (nonvolatile and volatile compounds), biological (origin, variety) and physical (physical integrity) (Guehi *et al*., 2008). Among nonvolatile flavor precursors, monosaccharides, disaccharides, oligosaccharides and some L-amino acids can contribute to the sweet taste of cocoa, while FA can contribute to acid taste. Tannin molecules like epicatechins, catechins, and procyanidins (total polyphenols) can confer bitterness and astringency. Alkaloid molecules (methylxanthines) can also contribute to a bitter cocoa taste (Jinap, Thien, & Yap, 1994). Thus they condition the sweetness, bitterness, acidity and astringency of cocoa and its derivatives.

Volatile compounds appear in cocoa post-harvest stages, such as fermentation and drying. These steps occur in the origins of cocoa beans by generating heterogeneous materials. As previously mentioned, variety and physical integrity (that depend on postharvest practices) are important factors for volatiles to form.

During fermentation and roasting, pleasant volatiles that determine chocolate odor, (such as aldehydes, ketones and pyrazines) are formed. Jointly with esters and alcohols, these compounds are also related with sweet odor (Rodriguez-Campos *et al.*, 2012). Properly dried beans usually have a long shelf life, a crisp texture and plump appearance, a well-oxidized interior and good flavor without excessive acidity, hammy, smokiness or other off-flavor notes (Jinap *et al*, 1994).

In contrast, inappropriate post-harvest handling (e.g. amount of mucilage in pods) can generate high contents of volatile fatty acids (VFA) like acetic, propionic, butyric, isobutyric, isovaleric acids (C2-C5), which cause strong acidic flavors and off odors. These off odors include rancidity, musty, stale, cheese rind, unpleasant and hammy flavors (García-Alamilla *et al*., 2007; Vázquez-Ovando, Chacón-

Introduction

Martínez, Betancur-Ancona, Escalona-Buendía, & Salvador-Figueroa, 2015). Nevertheless, VFA can decrease during roasting (Jinap *et al*., 1994).

The free amino acids, short-chain peptides and reducing sugars formed during the fermentation process can also contribute to cocoa flavor development during roasting in Maillard reactions. Aldehydes and pyrazines are produced as a result of this reaction. Tetramethylpyrazines (TMP) reach their maximum level upon medium roasting; trimethylpyrazines (TrMP) increase steadily throughout the roasting process and 2,5-dimethylpyrazines (DMP) rise under strong roasting conditions. The sensory evaluation shows that a normal roasting degree is linked to high concentration ratios of TMP/DMP and TMP/TrMP between about 1.5 and 2.5, respectively. Low values for the above ratios are linked to over-roasted cocoa beans (Aprotosoaie, Luca, & Miron, 2016). So they contribute to high quality chocolates, and these molecules are desirable in cocoa beans (Afoakwa, Paterson, Fowler, & Ryan, 2009). A more extensive description can be found in (Aprotosoaie, Luca, & Miron, 2016).

Regarding the analysis of aroma and flavor compounds, on the one hand, part of the aroma analysis is done by determining the aroma precursors that are free amino acids, oligopeptides, and reducing sugars. The analyses of these compounds have been previously described. This section reports only the methods used to study the combination between aroma precursors and sensory attributes.

A profounder understanding of the aroma profile can be attained by the determination of individual aromatic compounds. The determination of aroma compounds is usually made by their extraction, separation and detection. Studies have been published using different extraction methods, such as headspace-solid phase microextraction (HS-SPME) (Miriam Torres-Moreno, Tarrega, & Blanch, 2014), solid-phase microextraction (SPME) (Humston, Knowles, McShea, & Synovec, 2010), aroma extraction and dilution analyses (AEDA) and solvent-assisted flavor evaporation (SAFE distillation) (Chetschik *et al*., 2018).

To separate compounds, gas chromatography GC is frequently used. Flame ionization detector FID (Cambrai *et al*., 2010), mass spectrometry (MS) or, for more accurate detection times, fly mass spectrometry (TOFMS) (Humston., 2010) are used for detection purposes.

Many studies about the determination and/or changes produced in the flavor, aroma and taste during cocoa fermentation (Crafack *et al*., 2014), roasting (Torres-Moreno *et al*., 2014), between different cocoa clones or varieties (Liu *et al*., 2017), and between different cocoa products (Chetschik *et al*., 2018), have been recently published. Torres-Moreno *et al*., (2014) extracted volatile compounds from dark chocolate using HS-SPME followed by GC-MS to determine the influence of the roasting process on chocolate aroma formation. Variations have been found in the chocolate aroma profile and their concentrations according to roasting time and geographical origin (Torres-Moreno., 2014).

Changes in the aroma of cocoa beans with moisture damage have been analyzed in cocoa beans of different origins (Costa Rica, Ghana, Ivory Coast, Venezuela, Ecuador and Panama). SPME sampling coupled to two-dimensional gas chromatography combined with time of fly mass spectrometry (GC×GC–TOFMS) has been applied for such assessments. Twenty-nine compounds have been detected as a result of moisture damage (Humston., 2010). Similarly, gas chromatography coupled to a flame ionization detector (FID) and MS has been used to distinguish different cocoa types and their derivatives (Cambrai *et al*., 2010).

Thanks to a high sensitivity, selectivity and reproducibility of HS-SPME-GC–MS, the method is being increasingly used in combination with chemometrics. This determination technique and principal components PCA have been used to simultaneously understand the behavior of several aroma components (Cambrai *et al*., 2010)**.** Li, *et al*., (2012) detected 80 volatile aroma compounds in cocoa powders of different degrees of alkalization by the aforementioned GC-MS technique. Among these compounds, a high acetic acid concentration was determined. Moreover, a decreasing trend of this acid while increasing the degree of alkalization was reported

(Li *et al*., 2012). HS-SPME-GC-MS has also been used to evaluate the inoculation effect of starter cultures and fermentation techniques on the volatile aroma and sensory profile of chocolate. As a result, 56 volatile chocolate compounds have been identified and aromatic profiling differences have been linked to fermentation technique types, but not to the used starter cultures. However, the differences were too small to change consumer perceptions (Crafack *et al*., 2014).

Other aroma extraction methods include the aroma extraction and dilution analyses (AEDA) and solvent-assisted flavor evaporation (SAFE distillation), and both can be coupled to GC-MS. Chetschik *et al*. (2018) used the SAFE method to characterize the aromas of cocoa pulp, and how they are transformed during fermentation. These authors found higher 2-phenylethanol and 3-methylbutyl acetate concentrations in cocoa pulp than in cocoa beans in several fermentation stages. Conversely, quantities of odorants, such as linalool and 2-methoxyphenol, have been observed at larger concentrations in cocoa beans (Chetschik *et al*., 2018).

In another study (Van Durme, Ingels, & De Winne, 2016), the authors used the in-line roasting hyphenated with a cooled injection system coupled to a gas chromatograph–mass spectrometer (ILR-CIS-GC–MS) to assess fermentation quality and the overall potential formation of cocoa aroma. For this purpose, data on unroasted cocoa were compared with data on conventional roasted cocoa beans obtained by headspace solid phase microextraction (HS-SPME-GC–MS). The results of this analysis revealed that similar formation trends of important cocoa aroma markers were found according to fermentation quality. These main markers of cocoa aroma were aldehyde, pyrazines, aldehydes (amyl alcohols), and pyrazines tetramethylpyrazine (TMP) and trimethylpyrazine (TrMP), which are present at high concentrations when cocoa beans are well-fermented. The aforementioned method requires no sample preparation and can be performed in short times (<1 h).

Apart from methods based on the separation and identification of compounds, new innovative, faster and robust analytical techniques to determine aromatic compounds are being proposed. Concretely, the hyphenated HS-SPME-MS-nose

configuration, based on mass fingerprinting and pattern recognition, uses the hyphenated dynamic headspace-chemical sensor configuration. This equipment has a fully automated sample preparation unit for the online dynamic headspace isolation of cocoa aroma compounds. This technique has been used for the differentiation by the origin and fermentation degree of roasted fermented cocoa beans (from Indonesia, Peru, Ghana and Vietnam) by a hierarchical cluster analysis (HCA), PCA and one classification algorithm, namely soft independent modeling of class analogy (SIMCA). So a clear separation of fine flavor cocoa variety Criollo was possible, as was classifying samples according to their degree of roasting (Diem *et al.*, 2015).

Regarding origin, Liu *et al.* (2017) made a comparison of the aroma compounds of cocoa liquors from Asia, Africa and Oceania by gas chromatography-olfactometry-mass spectrometry (GC-O-MS). With this study, components at high concentrations were found, such as 3-methylbutanal, acetic acid, tetramethylpyrazine, and 3-methylbutanoic acid, and a relation between the aroma profile and origin was found by PCA (Liu *et al.*, 2017).

2.8 Polycyclic aromatic hydrocarbons, toxins and heavy metals

Cocoa samples can also contain compounds that could be considered of risk for humans. These compounds can come from soil contamination (i.e. heavy metals (HM), or can be generated during manufacturing practices (i.e. polycyclic aromatic hydrocarbons (PAHs) and mycotoxins). The levels of some of these compounds are regulated by the European Food Safety Authority (EFSA) (European Commission, 2011). The methods normally used and the studies carried out to control their presence are described below.

2.8.1 Polycyclic aromatic hydrocarbons (PAHs)

PAHs can be generated during incomplete combustion and are widely present in the environment. These compounds can contaminate foodstuffs and are related to human toxicity (carcinogenic, genotoxic, mutagenic) (Cordella *et al*., 2012). As they are lipophilic, their determination is usually made in cocoa butter. A frequent way to analyze PAHs in cocoa samples is to extract them from the sample by the method based on the stirred saponification of 1 g of cocoa butter in KOH (1:6), 1M at 80°C.

After extraction, the determination is made by HPLC coupled to a fluorescence detector. Four PAHs have been mainly determined, namely benzo(a) anthracene, chrysene, fluoranthene, and benzo(b) pyrene (Bratinova, Karasek, Buttinger, & Wenzl, 2015). Sess-Tchotch *et al*., (2018) used the aforementioned extraction and determination method and found limits of detections (LoDs) ranging from 0.01 µg/kg to 0.033 µg /kg for these compounds (Sess-Tchotch *et al*., 2018).

Another example of the identification and quantification of polycyclic aromatic hydrocarbons in cocoa beans was recently presented by Belo *et al*., (2017). These authors used an accelerated solvent extraction before GC-MS to determine eight PAH in cocoa beans. The evaluation of the method was made by analyzing relative standard deviations (RSD) under repeatability and precision conditions, and average recoveries. The authors found precision with RSD ranging from 2.57% to 14.13% and from 4.36% to 19.77% under repeatability and intermediate precision conditions, respectively. The average recoveries of the eight PAH ranged from 74.99% to 109.73%. These parameters, limits and measurement uncertainties met the performance criteria set by EU regulations.

2.8.2 Toxins

Not many studies about toxins in cocoa and its products can be found. The few studies published to date show that the most widely studied toxins in cocoa and its

products are ochratoxin A (OTA) (Kutsanedzie *et al.*, 2018) and aflatoxins. Ochratoxin is a mycotoxin that is formed by species of *Aspergillus* and *Penicillium*. Aflatoxins are formed by *Aspergillus flavus*, *A parasiticus*, and other *Aspergillus* spp. The most important aflatoxin, due to its occurrence, is aflatoxin B1, which is classified as carcinogenic (Group 1). The presence of ochratoxins in cocoa can lead to such serious health problems that the European Commission has set a tolerable weekly intake (TWI) of 120 ng/kg b.w. However, no maximum limit has been set for cocoa and cocoa products as these products do not contribute significantly to OTA exposure in diet (European Commission, 2010). No maximum limits have been set for aflatoxin (Turcotte, Scott, & Tague, 2013).

The most widespread technique to analyze toxins in cocoa is HPLC. To analyze ochratoxin in cocoa powder, Brera, Grossi and Miraglia (2005) developed an HPLC method based on OTA extraction from samples by blending with an aqueous solution of bicarbonate, diluting with a solution of phosphate buffer saline, filtering and cleaning-up by an immunoaffinity column (IAC) that contained antibodies specific to OTA. After washing the immunoaffinity column, OTA was eluted with methanol, separated by reversed-phase HPLC and quantified by fluorescence detection. This method was validated by an interlaboratory study, and allows the detection and identification of different OTA within the 0.1-2 µg/kg range. The same method was followed with drinking chocolate and cocoa powder to also detect ochratoxin (Cubero-Leon, Bouten, Senyuva, & Stroka, 2017). In this study, the authors found that the mean recoveries ranged from 85% to 88%, the RSD values went from 13.7% to 30.7% and the resulting Horwitz ratios, according to the Horwitz function modified by Thompson, fell within the 0.6-1.4 range for cocoa and drinking chocolate, respectively.

In a recent study that aimed to determine toxins in different cocoa products, toxins extracts were cleaned by AflaOchra (IAC) columns before HPLC separation. Toxin detection was performed by a post-column photochemical reactor for aflatoxin B1 and G1 (due to derivatization) and by fluorescence for OTA. The method's limits of

quantification (LOQ) were 0.16 ng/g (OTA) and 0.07 ng/g (aflatoxin B1). The OTA levels in the different analyzed samples were 1.17 ng/g in natural cocoa, 1.06 ng/g in alkalized cocoa, 0.49ng/g in baking cocoa, 0.39ng/g in dark chocolate, 0.19 ng/g in milk chocolate and 0.43 ng/g in cocoa liquor. Regarding aflatoxin, the following incidences were found: 0.86 ng/g in natural cocoa, 0.37 ng/g alkalized in cocoa, 0.22 ng/g in baking chocolate, 0.19 ng/g in dark chocolate, 0.09 ng/g in milk chocolate and 0.43 ng/g in cocoa liquor (Turcotte *et al.*, 2013).

2.8.3 Heavy metals

Heavy metals (HM) are naturally present in foodstuffs. These compounds are toxic to humans. Cadmium (Cd) is a heavy metal present in several foods consumed daily and in larger quantities, including cocoa. In order to maintain and control the amount of Cd in the human diet, the European Commission has set maximum Cd limits in certain products (European Commission, 2006), for example 0.10 mg/kg in milk chocolate with < 30% total dry cocoa solids, 0.30 mg/kg in milk chocolate with ≥ 30% total dry cocoa solids or 0.60 mg/kg in cocoa powder sold to end consumers or as an ingredient in sweetened cocoa powder sold to end consumers (drinking chocolate). In this context, monitoring the presence of this and other HM in cocoa products is a growing necessity.

To ensure compliance with regulations, CODEX STAN 228 (2001) suggests some Cd analytical methods, such as atomic absorption spectrometry (AAS) after incineration or microwave digestion (using HNO_3) and Anodic Stripping Voltammetry (ASV), of which AAS is more widely used.

Such is the concern today about the presence of Cd in cocoa and derived products that many studies have been conducted in the last 5 years to determine the amount of Cd present in cocoa derivatives. Cd has been determined in cocoa beans (Chavez *et al.*, 2015) and plants from Ecuador (Chavez *et al.*, 2016), in cocoa trees and leaves from Peru (Arévalo-Gardini, Arévalo-Hernández, Baligar, & He, 2017);

in cocoa beans from Indonesia (Assa, Noor, Yunus, Misnawi, & Djide, 2018); in cocoa powders and chocolates in the USA (Abt, Fong Sam, Gray, & Robin, 2018), in raw cocoa and processed chocolate mass from Poland (Kruszewski, Obiedziński, & Kowalska, 2018), and in Italian cocoa powder and chocolate (Dico *et al*., 2018). In the study of Chavez *et al*., 2015, the authors determined Cd in cocoa plant materials (ground leaves, shells or beans). For their analysis, samples (ground leaf, shell or bean) were digested with nitric acid (HNO_3) (Jackson *et al*., 1986). The digested samples were diluted with distilled water and filtered through a membrane filter prior to the Cd analysis. Then the Cd concentrations in plant digesters were determined by inductively coupled plasma optical emission spectrometry (ICP-OES).

In another article, the Cd concentrations in cocoa beans from Indonesia were established by atomic absorption spectrometry after digesting samples with HNO_3 and H_2O in a microwave. The Cd concentration in these samples was below the LOD of 0.100 mg/kg (Assa, Noor, Yunus, Misnawi, & Djide, 2018). Finally, Abt *et al*., (2018) determined Cd content in cocoa powder and chocolate products on the US market, and concluded that the Cd contained in these products ranged from 0.004-3.15 mg/kg.

3. Other analytical methods for cocoa beans trading across the value chain

Apart from the compositional analysis, other quality control analyses are done before and during the commercialization of cocoa beans. This section indicates how these analyses are applied. Quality control begins in the place of origin. To do so, conventional methods to assess and control correct fermentation, size, and even the integrity, of beans are widely used after sampling the total batch (FCC, 2018; ICCO, 2018). For fermentation quality assessments, the standard method applied is the cut test. It involves counting 300 beans. These 300 beans are then cut lengthwise through the middle and examined to infer the physical (integrity, color) and sensorial

characteristics of cocoa fermented beans, which provides an indication of quality (Lee & Djauhari, 2013; Schwan, 1998). During the cut test, the number of defective cocoa beans can be assessed. These defects can consist of beans with mould, damage caused by insects, and germinated or flat beans. The results are expressed as a percentage of 300 beans examined per defect kind. The amount of defective beans determined through cut tests is an indication of flavor characteristics (ICCO, 2018). Bean size is established by counting the number of cocoa beans per 100 g. By considering this, they are classified into three grades as follows: grade 1 ($\leq$ 100 beans per 100 g), grade 2 (101–110 beans per 100 g) and grade 3 (111–120 beans per 100 g). In bean mass (or weight) terms, the standard states that bean cocoa mass should be at least 1.0 g (CAOBISCO-ECA-FCC, 2015).

Before commercialization, other control parameters can include color, pH and titratable acidity (Hinneh *et al.*, 2018). During cocoa transformation, cocoa shell determination after shelling is an important factor as it affects some final product characteristics, such as flavor or taste. It can also be responsible for off-flavors. The fiber content in cocoa shells is very high. Thus it can be a problem for the grinding process as it can cause equipment abrasion in some cases (Mendes & Lima, 2007; Quelal-Vásconez *et al.*, 2019). During shelling, cocoa shells (approx. 12-20% of the cocoa bean) cannot be completely removed. In order to guarantee that cocoa powders have been well peeled and not adulterated with cocoa shells, the Codex Alimentarius establishes that cocoa shells including germ must be present, but below 5% (in fat-free dry cocoa) (Codex Alimentarius, 2014; Okiyama *et al.*, 2017). The official methods followed to analyze cocoa shells content are methods AOAC 968.10 and 970.23 (Codex Alimentarius, 2014). The first method, called the spiral vessel count, consists of counting spiral vessels in a defatted, ground and digested sample with the help of a microscope adjusted to mold counting (field of view 1.382 mm at 100 x) (AOAC, 2006). The second method, called the stone cell count, consists of counting the stone cells present in samples assisted by a microscope after laborious preparation (AOAC, 1984).

As observed in Sections 2 and 3, all the conventional methods followed to determine cocoa components or quality control during cocoa bean trading focus on destructive determinations. The inability to use the analyzed raw material, in combination with very long analytical procedures, high solvents utilization and waste production, and the need for highly skilled operators, mean that fast nondestructive alternative technologies must be developed.

4. Fast nondestructive technologies applied in the cocoa industry

This section contains an exhaustive analysis of the nondestructive technologies applied in the cocoa industry. A review of the different studies carried out with nondestructive techniques, analyzed products, the equipment used for analyses, measurement parameters, traditional methods used as references, chemometric model calibration and validation details. The results are presented in Table 3.

4.1 Types of non-destructive techniques used

Several fast nondestructive technologies, such as electronic tongue, electronic nose, hyperspectral image, terahertz spectroscopy and infrared spectroscopy, have been widely explored and applied in the cocoa industry (Table 3). Electronic tongue has been used for the rapid identification of cocoa beans according to their geographical locations (Teye *et al*., 2014a). Electronig tongue and near infrared spectroscopy, together with a chemometric analysis, has been used for the accurate classification of cocoa bean varieties (Teye, Huang, Takrama, & Haiyang, 2014c) and for the rapid determination of total polyphenols contents in cocoa beans (Huang *et al*., 2014). An electronic nose / gas chromatography-mass spectrometry (GC-MS) system combined with artificial neural network (ANN) has been used for detrmining roasting degree in cocoa beans (Tan & Kerr, 2018). Electronic nose combined with pressure controlled generated stimulation has been used in chocolate classification

Introduction

(Valdez & Gutiérrez, 2016). The hyperspectral image analysis has been used for cocoa bean quality assessments (Soto *et al*., 2015) and to predict the fermentation index, polyphenol content and antioxidant activity in single cocoa beans (Caporaso *et al*., 2018). These analyses have been done with whole cocoa beans and spectra measurements have been correlated with conventional PLS determinations with promising results. Terahertz (THz) spectroscopy has been utilized to control tempering in chocolate factories (Weiller, Tanabe, & Oyama, 2018). Terahertz (THz spectroscopy energy corresponds to collective molecular macro-vibrations and is considered a promising potential to identify macromolecules (i.e., polymer and biomolecules). This nondestructive noncontact technique has been used to characterize polytypes of crystals formed on the basis of FA combination in the chocolate structure. For this purpose, two chocolates are analyzed and the measurements normally taken by X-ray diffraction (XRD) are compared with optical microscopic observations and THz spectroscopy measurements, with similar results (Weiller, *et al*., 2018). Infrared spectroscopy has been used to predict major (moisture, carbohydrate, fat, protein) or minor functional compounds (theobromine, catechin, organic acids, etc.) (Álvarez, Pérez, Cros, Lares, & Assemat, 2012; Huang *et al*., 2014; Krähmer *et al*., 2015; Veselá *et al*., 2007) and for quality control (discrimination of cocoa beans according to geographical origin, prediction of cocoa powder adulterations, prediction of methylxanthines and polyphenols in alkalized cocoa powder, etc.) (Quelal-Vásconez *et al* 2020; Quelal-Vásconez *et al*., 2019; Quelal-Vásconez, Pérez-Esteve, Arnau-Bonachera, Barat, & Talens, 2018; Teye, Huang, Dai, & Chen, 2013).

Of all of the above-described technologies, infrared spectroscopy offers a number of important advantages over traditional chemical methods. It is nondestructive, noninvasive, requires minimal or no sample preparation, its precision is high and it can act as a multi-analytical technique because several determinations can be simultaneously made. Infrared spectroscopy also offers the possibility of measuring physico-chemical properties (Veselá *et al*., 2007).

36

4.2 Infrared spectroscopy

Infrared spectroscopy (IR) involves the interaction of infrared radiation with matter. It is conventionally divided into three wavelength regions: near-infrared (NIR: 750–2500 nm or 13333–4000 cm^{-1}), mid-infrared (MIR: 2500–25 000 nm or 4000–400 cm^{-1}), and far-infrared (25–1000 µm or 400–10 cm^{-1}). The distinction made among these three regions may vary depending on the type of instrumentation used to acquire IR spectral information.

4.2.1 NIR spectra acquisition

The IR method or technique is run with an instrument called an infrared spectrometer (or spectrophotometer) to produce an infrared spectrum. A generalized spectrophotometer has four parts: 1) an energy source; 2) a wavelength selection device; 3) a detector; 4) a data processing system.

The most explored technologies for cocoa studies are near infrared spectroscopy (NIR), Fourier-transform near infrared spectroscopy (FTNIR) and, to a lesser extent, Fourier-transform infrared spectroscopy (FTIR). The term Fourier-transform infrared spectroscopy originates from the fact that a Fourier transform (a mathematical process) is required to convert raw data (collected in frequencies in an interferogram) into the actual spectrum. In an NIR instrument, values are reported in nm, generally from 900 to 2500 nm, or from 650 to 2500 nm if the visible region is

included (Nielsen, Snitkjaer, & Van Den Berg, 2008). The values with an FTIR instrument are generally reported in cm^{-1}, from 4000 to 10000 cm^{-1} (Teye & Huang, 2015a) or from 4000 to 12500 cm^{-1} (Sunoj, Igathinathane, & Visvanathan, 2016).

Several optical alternatives are available for IR spectroscopy: 'reflectance', 'transmittance', 'transflectance', and 'interactance' (Alander, Bochko, Martinkauppi, Saranwong, & Mantere, 2013; Cortés, Blasco, Aleixos, Cubero, & Talens, 2019). The majority of studies for cocoa powder (Quelal-Vásconez *et al*., 2018) or cocoa beans (Caporaso *et al*., 2018) use reflectance (Table 3), but

transflectance has been used for semi-solids and liquids like cocoa butter or chocolate (Bolliger, Zeng, & Windhab, 1999).

Table 3: Non-destructive research of cocoa and sub-products applied in off-line and in-line process mode

Analyzed parameters	Conventional	Equipment and conditions	# Samples or spectra	Chemometrics	Calibration and cross-validation	External test set	Author
Beans-Postharvest /Fermentation Total Fungi count	Cut test (Bean fermentation levels, flavor, color, fungi presence and other foreign materials)	FT-NIR (Antaris II model, Thermo Fisher Company in the U.S.A); wavelength range (WR): 1000-2500 nm; scans: 5, Transmittance	Total: 95 samples; Calibration:57 ; Prediction:38	PLS; (Si-PLS); (Si-GAPLS); (ACO-PLS) (CARS-PLS)	Correlation coefficient of calibration: 0.97 RMSEcv: 0.402	Correlation coefficient of prediction: Rp 0.951 RMSEP (CFU/mL) 0.384 RPD: 4.32	Kutsanedzie et al., (2018)
Beans-Postharvest/ Compositional analysis Fermentation index (FI), pH, and total polyphenol content (TPC)	FI, pH, TPC	FT-NIR spectroscopy system (MATRIX-I, Bruker optics, Germany) using integrating sphere; WR: 12500-3600 cm^{-1} or 800 to 2778 nm; resolution 8 cm^{-1}; scans: 64	72 spectra (24 samples 3 replications)	PLS	FI: R^2 of 0.88; RMSEcv=0.06; RPD: 2.74; pH: R^2 of 0.76; RMSEcv=0.26; RPD: 2.05; Total TP:R^2 of 0.84; RMSEcv=0.93; RPD: 2.53; 0.535 and 1.242, TPC 6.48 and 15.58 mg g^{-1}	NA	Sunoj et al., (2016)

Table 3 (continued 2)

Analyzed parameters	Conventional	Equipment and conditions	# Samples or spectra	Chemometrics	Calibration and cross-validation	External test set	Author
Beans-Postharvest/ Compositional analysis -Phenolic substances, organic acids, epicatechin, lactic acid, fermentation time, pH	High performance liquid chromatography HPLC, FI, sensory test (QDA), Gas chromatography mass spectrometry GCMS, cut test	FT-NIR (Multi-Purpose Analyzer, Bruker Optics, Ettlingen, Germany), WR: 12500 to 3600 cm⁻¹; resolution 8 cm⁻¹; scans 32	48 samples	PLS	Leave 10% out Theobromine 1.73-3.02 mg/100g	NA	Krähmer *et al.,* (2015)
Beans-Postharvest/ Fermentation - Fermented and unfermented	HPLC, FI, sensory test, GCMS, cut test	FTNIR	Classification: FC=26; UFC=26; Adulterated 80. Calibration 90; Prediction 42	SVM, SiPLS	Leave one out	100% classification; RMSEP:0.98, prediction since 5%	Teye, Huang, Lei, & Dai, (2014b)
Beans-Postharvest -TPC	Colorimetry (Folin-Ciocalteu), Thin layer chromatography and HPLC	ET: Data collection: Astree brand (Alpha MOS Company, Toulouse, France), potentiometric chemical sensors. FT-NIR: Antaris II FT NIR (Thermo Electron Company, USA) equipped with an indium gallium arsenide (InGaAs) photodiode detector. WR: 10000-4000cm⁻¹; Scans 32; interval 3.85 cm⁻¹; resolution 8 cm⁻¹, 25ºC; humidity 60%	110 samples (80 calibration and 30 prediction)	Fusion techniques: low level of abstraction with PCA, (Si-PLS)	NA	Optimal data fusion model: Rp 0.982, RMSEP 0.900 g g⁻¹ and bias 0.013 in the prediction set	Huang *et al.,* (2014)

Table 3 (continued 3)

Analyzed parameters	Conventional	Equipment and conditions	# Samples or spectra	Chemometric techniques	Calibration and crossvalidation	External test set	Author
Beans, ground and sieved 400 µm mesh - Postharvest/ Fermentation - pH and FI	pH and FI	FT-NIR Antaris II (Thermo Electron Company, USA) with an integrating sphere, WR: 10000-4000cm⁻¹; scans 32; interval 3.85 cm⁻¹, resolution 8 cm⁻¹, 25ºC	Categories Fermented 80, partly fermented 25 and unfermented 25. Model 90 calibration, 40 validation	(BPANN) ;(Si-BPANNR)	Leave one cross-validation (LOO-CV)	pH was Rp = 0.98 and RMSEP = 0.06, while for FI was Rp = 0.98 and RMSEP = 0.05	Teye et $al.$, (2015b)
Cocoa beans, cocoa liquor - Postharvest/ Fermentation - Amount of ammonia nitrogen (NH3)	Conway technic (Conway & Byrne, 1933)	NIR FOSS 6500 monochromator (Foss, Silver Spring, MD) using a spin cell sample module; WR: 400-2500 nm; intervals 2 nm; scans 32. Reflectance	190 samples and spectra	PCA (Mahalanobish distance (H) no higher than 3), PLS	R^2c= 0.975, SEC16 ppm, SE$_{CV}$ 24 ppm; Range 25-441 ppm	R^2p 0.935, SEP 20 ppm Range 46-332 ppm	Hue et $al.$, (2014)
Beans, ground and sieved 500 um mesh - Postharvest/ Compositional - Fat	Fat (Soxhlet extraction apparatus - SOEP, Microwave-assisted process-MAP)	FT-NIR Antaris II (Thermo Electron Company, USA) equipped with an InGaAs photodiode detector. WR: 10000-4000cm⁻¹; scans 32; interval 3.85 cm⁻¹, resolution 8 cm⁻¹, 25ºC	80 (50 calibration, 30 prediction)	(Si-PLS), SVMR	Leave one out cross-validation (LOO-CV)	Si-SVMR_ RMSEP=0.015 and Rp=0.9708	Teye & Huang, (2015a)

Table 3 (continued 4)

Analyzed parameters	Conventional	Equipment and conditions	# Samples or spectra	Chemometric techniques	Calibration and crossvalidation	External test set	Author
Beans - Postharvest/ Drying - Compositional analysis	AOAC Method to determine purines and procyanidins by HPLC	NIR: Monochromator (model 6500; Foss NIRSystems, Laurel, Maryland; WR 400-2500 nm; intervals 2 nm, scans 32, Reflection	Wet chemistry: 40 NIR: 470(Fat), 342 (caffeine), 343(theobromine), 224 (epicatechin)	Neighbourhood Mahalanobish Distances PLS	25% of the samples randomly selected leave out, 4 times.	NA	Álvarez et al., (2012)
Beans - Postharvest/ Fermentation - Microbial profiles	Cut test (visual)	NIR System 6500, Inc. USA, 400-2500, silicon detector, WR: 400-1100 and lead sulphide detector, WR: 1100- 2500; intervals 2nm, 45º, scans 16, Reflectance	50 kg	(DGGE) PCA, PLS 2	Correlations 0.87 (bacterial derived DGCE spectra) and 0.81 (yeast derived DGCE spectra)	NA	Nielsen et al., (2008)
Beans - Postharvest/ Fermentation - FI, TPC and antioxidant activity (AA)	HPLC ABTS (Antioxidant activity); QUENCHER "Quick, Easy, New, Cheap and Reproducible method"	HSI (Hyperspectral image) system Gilden Photonics Ltd. (Glasgow, UK) SWIR camera (Specim, Oulu, Finland) with a cooled 14 bit 320x256 pixel HgCdTe detector; WR; 1000-2495 nm; resolution 6 nm	17 beans 170 beans/batch; 170 samples; 340 spectra; Calibration:240; Prediction:100	PLS	Full cross-validation, FI: R^2=0.57 (RMSEc=0.22; RMSEcv=0.24) TP: R^2=0.82 (RMSEc = 23.35mg ferulic acid g^{-1}, RMSEcv= 28.09; AA:R^2=0.76 (RMSEc =55.25 mmol Trolox. kg^{-1}, RMSEcv =59.23	FI: R^2 = 0.50 (RMSEP = 0.27, RPD = 1.40), TP: R^2=0.70 (RMSEP= 34.1 mg ferulic acid g^{-1}, RPD = 1.77) and AA: R^2= 0.74 (60.0 mmol Trolox . kg^{-1} , RPD = 1.91)	Caporaso et al., (2018)

Table 3 (continued 5)

Analyzed parameters	Conventional	Equipment and conditions	# Samples or spectra	Chemometrics	Calibration and crossvalidation	External test set	Author
Cocoa butter -Sensorial -Fatty acids -Crystal structure of chocolate	X-ray diffraction (XRD) (characterization of crystal structure)	THz spectroscopy 1064 bn Q-switch pulsed Nd: YAG Laser as pump for Cr: For sterite lasers	2 chocolates, 3 samples prepared	NA	NA	NA	Weiller et al., (2018)
Cocoa powder - Industrial processing/ compositional analysis -Fat, nitrogen, and moisture	Soxhlet extraction (Fat), Nitrogen (Kjeldahl), Moisture (in platinum dish in an aerated oven at 100ºC)	FTIR: Bruker IFS-55 (Bruker, Germany) with single reflection ATR accessory, diamond cell (Golden Gate), DTGS detector and triangular apodization function Resolution 8 cm⁻¹; Background spectrum against the air; co-added scans 128; WR: 1100-2500nm and 4000-600 cm⁻¹. NIR: NIR Systems 6500 Persptop Analytical Company, USA. Tugsten filament lamp; resolution 2nm; 3 replicated (36 co-added scans). Ceramic used as photometric standard; scanning speed 62 s; Internal reference was the background spectrum	100	NIR-FTIR data fusion- outer product matrix for two spectra (vectors), PLS	NIR: $RMSE_{CV}$ = 7.0%, R^2=0.96 for fat, 1.7%, R^2 =0.98 for nitrogen, and 5.2%, R^2=0.94 for moisture. FTIR: $RMSE_{CV}$ = 10.4%, R^2 =0.94 for fat and 3.9%, R^2 =0.95 for nitrogen.	RMSEP (%): 1.2 fat, 0.10 nitrogen, 0.40 moisture.	Veselá et al., (2007)
Cocoa powder - Industrial processing/ compositional analysis -Moisture, fat, sucrose	Individual determination	NIR; WR: 1900-2320 nm single beam NIR spectrophotometer with three tilting filters (Pacific Scientific, Gardner Neotec Division, Model Compscan 3000), Reflection	calibration 65 samples, 10 samples prediction	PLS	R^2c=0.978, SEC 0.157; R^2c=0.987, SEC 0.100; R^2c= 0.998, SEC 0.526	SEP (%) Moisture 0.034, fat 0.051, sucrose 0.68	Permanyer & Perez, (1989)

Table 3 (continued 6)

Analyzed parameters	Conventional	Equipment and conditions	# Samples or spectra	Chemometrics	Calibration and crossvalidation	External test set	Author
Cocoa butter chocolate - Industrial processing/ physical properties – Cocoa butter, Viscosity, enthalpy (crystal content), and slope values. For precrystallized chocolate, analytical values such as viscosity and slope values – detected off-line and used for calibration of NIR spectroscopy.	Rheology, Viscosity (viscometer_Searle principle), Calorimetry (temper curves)	NIR: Universal spectrometer (NIRVIS with the software version BCAP 4.40 CH, Bühler AG, Uzwil, Switzerland); WR: 1000-2500 nm. Light fiber probe inserted in the outlet tube of the shear crystallizer through a special adapter. Cocoa butter: Transflection (transmission and reflection measurement used to transparent media) probe, measuring gap width 5mm (end of the probe and a standardized reflection surface) Chocolate: reflection probe	In line	PCA, PLS	Correlations with measurements of viscosity, crystal content (R = 0.975) and slope (R=0.945). Precrystallized chocolate correlation (R = 0.973)	NA	Bolliger et *al.,* (1999)

NA: Not apply, NE: Not specify

4.2.2 Multivariate data analysis

Due to the complex and the large amount of hidden information in IR spectral data, particular attention should be paid to data mining with chemometrics for the IR spectroscopy analysis. Multivariate data analysis or chemometrics is the science of extracting information from chemical systems by data-driven means. It offers lots of applications and diverse natures. Specifically, it is used in IR applications to extract rich information from IR spectra, including preprocessing spectral data, reducing variables, building calibration models (quantitative) and/or classification (qualitative) analyses, and model transfer, and all this to acquire more information from data (Martens *et al.*, 2003).

A multivariate analysis generally involves the following steps: data exploration, data preprocessing, quantitative or qualitative model calibrations, and finally external validation. Data exploration allows a group of samples to be found, the relation between variables, management with outliers samples by means of a PCA or a parallel factor analysis (PARAFAC) (Bro, 1997; Rodrigues, Condino, Pinheiro, & Nunes, 2016). Data preprocessing can be handled with preprocessing algorithms, such as: smoothing methods (Savitzky-Golay, Gaussian filter, median filter, moving average); normalization and scaling, detrending (Levasseur-garcia, 2018), 1st Derivate, 2nd Derivate-Savitzky Golay (Savitzky & Golay, 1964), Standard Normal Variation (SNV) (Teye, Uhomoibhi, & Wang, 2016), Orthogonal Signal Correction (OSC) (Wold, Antti, Lindgren, & Öhman, 1998) and Multiple Scatter Correction (MSC) to build and enhance calibration models (Su & Sun, 2017). The selected preprocessing method can be related to data features to, for example, rid up multiplicative and additive effects in spectra. As seen in Table 3, datasets are usually divided into calibration and validation, except

those carried out by Krämer *et al*., (2015) and Sunoj *et al*., (2018), who performed only cross-validation. Calibration datasets are composed of a different number of samples, from 65 (Permanyer & Perez, 1989) samples in the calibration set to 190 at the time of this study (Caporaso, Whitworth, Fowler, & Fisk, 2018). High accuracy has been obtained for calibration models by employing proper multivariate linear regressions, such as PLSR, PCR, SVMR, and other statistical algorithms like artificial neural networks (ANN) (Teye & Huang, 2015a; Teye, Huang, Lei, & Dai, 2014b; Teye *et al*., 2015b). PLS with variable selection, such as Sinergy Interval-PLS (Si-PLS), Ant Colony Optimization-PLS (ACO-LS), Competitive Adaptive Reweighted Sampling - PLS (CARS - PLS), Synergy Interval-Genetic Algorithm-PLS (Si-GAPLS) (Kutsanedzie *et al.,* 2018), Modified Partial Least Squares (mPLS) and Synergy Interval Backpropagation Neural Networks Regression (Si-BPANNR). Efficient classification results have been obtained with tools like support vector machine (SVM), discriminant partial least squares (PLS-DA) (Berrueta, Alonso, & Héberger, 2007), LDA, SIMCA, SVM, QDA and KNN (Teye, Uhomoibhi, & Wang, 2016; Teye, Huang, Han, & Botchway, 2014a). To build calibration models, all the spectra can be used, or variable selection methods also are employed to obtain computationally efficient algorithms.

However, a variable selection can be performed to avoid complex models. Table 3 also shows that full cross-valitadion is widely used during model calibration. The evaluation of model performance is made by parameters, such as the coefficient of determination of calibration, cross-validation and validation (R^2), coefficient of correlations (R), root mean error of calibration, cross-validation and validation and the relation deviation prediction (RPD). Sometimes both bias and slope are considered.

4.3 Applications

4.3.1 Non-destructive determination of constituents and industrial processing monitoring

Very few studies done with nondestructive technologies have been applied in the cocoa industry. Of these, the most frequently used non-destructive are NIR and FT-NIR (Table 3). The majority of studies have been done in the postharvest (fermentation/drying) stage of cocoa beans. Biochemical parameters like fat (Álvarez *et al*., 2012; Weiller *et al*., 2018), sugars, polyphenols, procyanidins (Whitacre *et al*., 2003), methylxanthines, moisture and pH (Krähmer *et al*., 2013; Sunoj *et al*., 2016; Veselá *et al*., 2007) have been evaluated. The aim of these studies was the quality control of end products, and/or the determination of authenticity through compositional analyses or by clustering samples from their spectral fingerprint (origin, varietal classification).

Near infrared light is sensitive to the sample's physical properties. These physical conditions can cause variations in measured spectra, and have been identified in spectra as multiplicative and additive effects. These effects, due to light scatter, are minimized using a sample of a small homogenized particle size (Barbin *et al*., 2018). Most studies have employed ground beans more than whole beans, partly as a way to minimize the aforementioned variations and effects (Barbin *et al*., 2018) (Table 3)

In relation to measurement modes, Dickens, (1999) defined four ways to implement measurement equipment into processes: (i) off-line: a sample analysis run away from the production line (i.e., laboratory); (ii) at-line: manual random sample extraction from the production line and an analysis performed close to the process line; (iii) on-line: samples separated from the

production line which, after being analyzed in a recirculation loop (by-pass), are returned. (iv) in-line: samples are analyzed on the running production line (in situ) (Dickens & Dickens, 1999; Osborne, 2000). Table 3 shows that the performance of this nondestructive analysis done in the off-line mode in almost all the studies carried out by NIR in cocoa beans. Only Bolliger performed an in-line application of NIR in 1999 to monitor the rheological properties (viscosity, melting enthalpy) of chocolate in the tempering stage.

In connection with cocoa bean fermentation, the degree of fermentation and flavor profile are routinely determined in both the trade and industry by a cut test (color check). Both assessments require specially trained personnel. Sensory evaluation is highly subjective depending on the sensory panel (Afoakwa *et al*., 2013). So fermentation has been the subject of different approaches, such as characterization by spectroscopic and chromatographic methods (Aculey *et al*., 2010). Accordingly, Table 3 shows that the lower value predicted by NIR is in ppm units of a metabolite (NH3) product of fermentation. NH3 contents have been found to fall within a range of 46-332 ppm with a standard error of prediction (SEP) of 20 ppm (Hue *et al*., 2014).

The fermentation of cocoa beans has been analyzed by NIR and EDGE to gain a better understanding of the fermentation mechanisms related to the microbiological factor. A good correlation between both measurements has been found (Nielsen *et al*., 2008). NIR integrated with an electronic tongue (ET) and multivariate analyses have been applied to perform a 100% (accuracy) classification of five cocoa bean varieties. Accurate classifications can be attributed to three functional groups (second overtone) of methylene (–CH_2), methyl (CH_3) and ethenyl (–CH=CH–). Theobromine, for instance, has one methyl group, while caffeine has two methyl groups. These compounds may play an important role in discriminating employed cocoa bean varieties (Teye, Huang, Takrama, *et al*., 2014c).

Bacteria (e.g. Staphylococcus aureus, Bacillus cereus) in cocoa powders have been found to affect their quality grades, and these bacteria can be detected by the FT-MIR spectral system (Ramalingam *et al*., 2009). The total fungi count (TFC) in cocoa beans has been evaluated by Fourier transform near infrared spectroscopy (FT-NIRS) combined with synergy interval-genetic algorithm-PLS (Si-GAPLS). This technique allowed a prediction coefficient of 0.975 to be obtained, along with a root mean square error of prediction (RMSE) of 0.384 CFU/mL and a ratio prediction deviation RPD of 4.32 (Kutsanedzie *et al*., 2018).

4.3.2 Authenticity and adulterations

Table 4 shows details of the studies carried out by the nondestructive techniques used to assess the authenticity and adulteration of cocoa products. Trilcová *et al*., 2004 showed that NIR and FTIR spectroscopy can be used as a very fast and reliable tool for cocoa powder authentication. The term authenticity refers to the inherent quality attributes of cocoa, and has been included in new food fraud authenticity policies and identified as product integrity (Manning, 2016). The authenticity of cocoa and its derivatives is determined by studies that aim to identify the origin of raw material, varietal purity, compositional parameters, detection of adulterants, etc.

The sensory characteristics of cocoa products have created an increasing consumer trend to choose cocoa of a specific origin. These preferences have allowed more appreciated cocoa origins whose quality is differenced by market prices. This differentiation has yielded bad commercial practices, like mixing more expensive cocoa beans of the highest quality and an outstanding origin with other lower quality cocoa kinds that are cheaper to obtain fraudulent economic benefits (Magagna *et al*., 2017).

Introduction

The determination of origin has been evaluated by the phenolic fingerprint (D'Souza *et al*., 2017). Most of these studies have been conducted by the compositional analysis mentioned in Section 2. Nondestructive technologies like NIRS have been applied to classify cocoa by its origin. As a result, classification percentages according to the geographical origin of cocoa beans of 90.63 (LDA), 75 (KNN), 96.88 (BPANN) and 100 (SVM) have been obtained by Teye *et al*., 2013.

Cocoa products and derivative fraud are related to intentional contamination, and to noncompliance to product descriptions and adulterations. The used adulterants are low-cost raw material (Van Ruth, Huisman, & Luning, 2017), such as different flours like carob or chicory, which have been processed to substitute cocoa powder (Loullis & Pinakoulaki, 2018; Salem & Ohaad Fahad, 2012). The NIR technique and the multivariate analysis have been used for the quantitative and qualitative detection of carob flour added to cocoa powder (Quelal-Vásconez *et al*., 2018). In another study, Quelal-Vásconez *et al*., 2019 quantitatively determined the presence of cocoa shells by NIR and a PLS model. These authors also classified between two categories of cocoa blends with 92.5% accuracy: (1) presence of < 5% cocoa shells; (2) presence of > 5% cocoa shells in cocoa powders.

Another adulteration type is to add different vegetal or animal fats to cocoa butter (Jahurul *et al*., 2018; Kucha, Liu, & Ngadi, 2018). These fats can come from pork, palm, Garnicia indica, Madhuca butyracea and other vegetable origins with lower market values (Reddy & Prabhakar, 1994). These fats are considered cocoa butter equivalents (CBE) and should not exceed 5% of the final cocoa product (EU, 2000). However, these less expensive materials and their intentional additions aim to lower production costs in industry.

50

No specific regulation exists for the products used as raw materials for the food industry. Industries (beverages, bakery, pastries) are responsible for testing their raw materials and for searching ways to detect adulterants (Beulens, Broens, Folstar, & Hofstede, 2005; Trafialek & Kolanowski, 2017) to ensure the precedence and content of these raw materials. Traceability is one of the ways to ensure the food safety of end products. Other studies have been done to detect added molecules, which are not declared in products like vanillin and ethyl vanillin (Pérez-Esteve *et al.*, 2016). Cocoa powder adulteration by means of identifying the fingerprints of cocoa powder polysaccharides has been studied, and has provided the possibility of finding from 15%, or more, cocoa shell powder and 10% exogenous plant material (Yang *et al.*, 2015).

FTIR has been applied to detect cocoa butter equivalents CBE (allowed in chocolate up to 5%: palm oil, illipe, sal, shea, kokum gurgi and mango kernel). FTIR is considered a rapid screening method to distinguish pure and vegetable fats, but a single global statistical model to predict the precise level of added fat is still not available. The large uncertainty in predicting CBE has been connected to the wide natural variability of samples (precise geographical origin). So it was difficult to detect CBE in CB mixtures (e.g. illipe) (Whitacre *et al.*, 2003)

Table 4: Authenticity evaluated with non-destructive methods

Analyzed parameters	Conventional	Equipment and conditions	# Samples or spectra	Chemometrics	Calibration and crossvalidation	Validation test	Author
Beans - Postharvest/Origin -Cocoa variety Moisture Ash Protein Fat Carbohydrates L* a* b* PH	Protein (total organic nitrogen- Kjeldahl procedure). Fat content (Soxhlet apparatus). Moisture (gravimetric method - drying 2 g of sample at 105 ºC to constant weight). Ash (muffle furnace at 550 ºC for 12 h) TPC (difference of components)	NIR Near-Infrared (XDS) model XM 1100 series-Rapid Content Analyzer Foss NIR Systems, Denmark; wavelength range WR: 400-2498 nm; 2nm intervals. Reflection	80 samples	PLS, SVM, LDA	Full cross-validation (leave one out)	The performance of SVM model was superior to LDA model, achieving an identification rate of 100%.	Barbin *et al.,* (2018)
Beans - Postharvest/Varietal discrimination/Authentication - Varietal discrimination	Physical characteristics and Cut test	Handled Raman spectrometer; Raman spectrometer with a 1064 nm Nd:YAG laser; Total exposure time for a Raman spectrum: 15 s; laser power set at 250 mW; resolution 4 cm^{-1}; 100 scans	20 samples (7 national and 13 CN-51 from different locations-25 samples per location were measured)	SVM	Leave one out	NA	Popp *et al.,* (2016)

Table 4 (Continued 2)

Analyzed parameters	Conventional	Equipment and conditions	# Samples or spectra	Chemometrics	Calibration and crossvalidation	Validation test	Author
Beans -Postharvest/-Varietal discrimination/Authentication	Liquid chromatography LC, gas chromatography GC, capillary electrophoresis, sensory evaluation or plasma atomic emission	ET: Data collection: Astree brand (Alpha MOS Company, Toulouse, France), potentiometric chemical sensors. NIR: Antaris II FT NIR (Thermo Electron Company, USA) equipped with InGaAs photodiode detector. WR: 10000-4000cm⁻¹. 32 scans, 3.85 cm⁻¹ interval, 8 cm⁻¹ resolution. 25ºC, 60% humidity, 9500-7500 cm⁻¹	65 calibration, 35 validation (samples)	Data fusion by normalization, PCA, SVM	NE	(NIRS and ET) has a classification rate between 83 and 93%. Data fusion (ET-NIRS) had a classification rate of 100%	Teye, Huang, Takrama, & Haiyang, (2014c)
Beans grounded -Geographical origin/Adulteration	Sensory evaluation GCMS, HPLC Colorimetry and inductively coupled plasma mass spectrometry	FT NIR Antaris II (Thermo Electron Company, USA) with an integrating sphere, WR: 10000-4000cm⁻¹). Rotating cup 120º, 32 scans, 3.85 cm⁻¹ interval, 8 cm⁻¹ resolution. 25ºC	194 samples, 130 calibration and 64 prediction	LDA, PLS-DA, KNN, BPANN, SVM	NA	Classification (%) LDA, KNN 90.63, 75, BPANN 96.88, SVM 100	Teye, Huang, Dai, & Chen, (2013)
Beans grounded, cocoa liquor -Postharvest/Industrial processing/ Authentication -Total procyanidin oligomers (monomer to decamer)	HPLC - Normal-phase separations of the procyanidin oligomers.	NIR Systems II 6500 (NIR Systems Inc. Silver Springs, Md., USA) WR: 400 to 2500 nm, 2 nm intervals, 20 scans. The samples were presented in an open cell (liquor measured to 50ºC). Reflection	96 samples	mPLS	Total procyanidins (mg/g) Mean prediction 9.89; SEC: 1.04; SECV: 1.09	10 samples prediction	Whitacre *et al.*, (2003)

53

Table 4 (Continued 3)

Analyzed parameters	Conventional	Equipment and conditions	# Samples or spectra	Chemometrics	Calibration and crossvalidation	Validation test	Author
Cocoa powder - Adulteration – cocoa shell	Blue value (colorimetric test), HPLC by detection of LGC (fatty acids of cocoa shell)	NIR FOSS 5000 (SILVER SRPING, MD, USA); WR: 1100-2500 nm; 2 nm intervals; 32 scans. Reflection	90 samples, 70 calibration and 20 prediction.	PLS, PLS-DA, variable selection	Full cross-validation or leave one out; R^2c=0.975; $RMSE_{CV}$:1.91;	R^2P:0.967; BIAS:0.195; RMSEP:2.43; RPD:5.03% accuracy:92.5	Quelal-Vásconez, et al, (2019)
Adulteration -Cocoa powder, carob flour	Chromatographic techniques	NIR FOSS 5000 (SILVER SRPING, MD, USA); WR: 1100-2500nm; 2 nm intervals; 32 scans. Reflection	12 cocoas 234 prepared samples	PLS, PLS -DA	Full cross-validation or leave one out; R^2c=0.98; $RMSE_{CV}$:2.9; SLOPE:0.981	Coefficient of determination for prediction (R^2) of 0.974 and a root mean square error of prediction (RMSEP) of 3.2%	Quelal-Vásconez, et al, (2018)
Butter (diluted 1:10, analytical grade acetone)- Industrial Processing/Authentication	Chromatography (triglycerides and fatty acids)	FTIR: Bruker IF28 FTIR spectrometer (Bruker Spectrospin Ltd., Coventry, United Kingdom) equipped with a mercury-cadmium telluride (MCT) detector cooled N_2; WR: 4000 to 600 cm^{-1}; Acquisition rate 20s^{-1}; Resolution 4 cm^{-1}; Improve signal-to-noise ratio 256 spectra/sample. Transmittance	192 samples (triplicate). 14 CB (10 pure- various geographical origins and 4 commercial mixtures) 18 CBE (12 mixtures and 6 pure CBE) and 154 mixtures of CB with CBE (5 to 20%). Training data encompass the full range under study. Training 186 spectra (62 in replicate) and 183 (61 in replicate)	PC-DFA r canonical varieties analysis (CVA), ANN, PLS	Good classification (10 and 20% adulteration level) of the training set	Non successful classification	Goodacre & Anklam, (2001)

NA: Not apply, NE: Not specify

54

Nondestructive technologies have been used to improve processes with new control and evaluation methods (e.g. the fermentation index, the degree of alkalization) and replaced or improving the conventional analysis methods (chromatography, sensory analysis, etc.). Several results about certain features like fat, moisture, color, proteins, pH (Moros, Iñón, Garrigues, & de la Guardia, 2007) and functional compounds (antioxidants) have been obtained by only spectra measurements. As the NIR technology has demonstrated its versatility, its applications are rapidly increasing not only to control the safety of cocoa products, but also to improve their quality, and to optimize times and costs.

Despite all the successful applications regarding the use of alternative methodologies to analyze and control the above-described cocoa quality, their implementation into the cocoa industry poses challenges, such us the simultaneous presence of a variety of chemical compounds (nutrients, phytochemicals, adulterants, contaminants, etc.) in cocoa products with diverse structures and concentrations. This circumstance makes spectrometric signals very complex and difficult to analyze. However, technology is rapidly advancing and new equipment include improved signal collection and software capable of performing chemometric analyses, which are key to acquire reliable information.

5. Conclusions

The analytical methods applied to control the quality and authenticity of cocoa products and their derivatives in industry and research laboratories have mainly been conventional ones to date. They are conventional because they have been used for years and are characterized by tasks like sampling, sample preparation to extract target compounds and quantitative determination by using chemical reagents. The majority of these methods are standardized and

used especially for monitoring and optimizing the process during cocoa flow production by individual analyses of attributes (color, pH, acidity and proximal analysis) by wet chemistry. Most of the advances made in these methods are related to analyte extraction to improve sensitivity accuracy and analysis speed, also to the application of multivariate data analyses. For sensitivity and accuracy determinations, chromatographic methods like HPLC and GCMS, and its inline utilization, are the most well-developed ones, while multivariate data analyses are mainly employed to determine the authenticity parameters (i.e. origin or varietal features) of cocoa products. The most explored nondestructive technique is spectroscopy, which is conducted within the near infrared range, and also within the medium infrared range to a lesser extent. Most NIR and FTIR studies have been conducted in the postharvest stage of cocoa beans by analyzing biochemical parameters like fat, sugars, polyphenols, procyanidins, methylxanthines, moisture and pH, or for the purpose of assessing the authenticity of cocoa and its derivatives by identifying the origin of raw material, varietal purity, compositional parameters or the detection of adulterants.

Acknowledgments

References

Abdullahi, G., Muhamad, R., Dzolkhifli, O., & Sinniah, U. R. (2018). Analysis of quality retentions in cocoa beans exposed to solar heat treatment in cardboard solar heater box. *Cogent Food & Agriculture*, *4*(1), 1–12.

Abt, E., Fong Sam, J., Gray, P., & Robin, L. P. (2018). Cadmium and lead in cocoa powder and chocolate products in the US Market. *Food Additives and Contaminants: Part B Surveillance, 11*(2), 92–102.

Acierno, V., Alewijn, M., Zomer, P., & van Ruth, S. M. (2018). Making cocoa origin traceable: Fingerprints of chocolates using Flow Infusion - Electro Spray Ionization - Mass Spectrometry. *Food Control*, *85*, 245–252.

Aculey, P. C., Snitkjaer, P., Owusu, M., Bassompiere, M., Takrama, J., Nørgaard, L., Petersen, M., Nielsen, D. S. (2010). Ghanaian cocoa bean fermentation characterized by spectroscopic and chromatographic methods and chemometrics. *Journal of Food Science*, *75*(6), 300–307.

ADM Cocoa Manual. (2006). The De Zaan Cocoa Manual. Archer Daniels Midland Company BV. The Netherlands.

Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2009). Matrix effects on flavour volatiles release in dark chocolates varying in particle size distribution and fat content using GC-mass spectrometry and GC-olfactometry. *Food Chemistry*, *113*(1), 208–215.

Afoakwa, E. O., Quao, J., Takrama, J., Budu, A. S., & Saalia, F. K. (2013). Chemical composition and physical quality characteristics of Ghanaian cocoa beans as affected by pulp pre-conditioning and fermentation. *Journal of Food Science and Technology*, *50*(6), 1097–1105.

Alander, J. T., Bochko, V., Martinkauppi, B., Saranwong, S., & Mantere, T. (2013). A Review of Optical Nondestructive Visual and Near-Infrared Methods for Food Quality and Safety. *International Journal of Spectroscopy*, *2013*, 1–36.

Introduction

Álvarez, C., Pérez, E., Cros, E., Lares, M., & Assemat, S. (2012). The use of near infrared spectroscopy to determine the fat , caffeine , theobromine and sun-dried beans of Criollo cocoa, *315*(August 2011), 307–315.

AOAC (1984). AOAC, Official methods of analysis (14th ed.). Association of Official Analytical Chemists (Arlington, Vol. A).

AOAC (1990b), International Fatty acids in oils and fats, preparation of methyl esters. AOAC official method 969.33, Official methods of analysis (15th ed.), AOAC International, Arlington, VA, 1990.

AOAC. (2006). AOAC Official Method 968.10 Spiral Vessel Count of Cacao Products, 2-3.

APEDA. (2015). (Agricultural and Processed Food Products Export Development Authority). Retrieved from http://agriexchange.apeda.gov.in/indexp/exportstatement.aspx. [Consulted 2019 January]

Aprotosoaie, A. C., Luca, S. V., & Miron, A. (2016). Flavor Chemistry of Cocoa and Cocoa Products-An Overview. *Comprehensive Reviews in Food Science and Food Safety*, *15*(1), 73–91.

Arévalo-Gardini, E., Arévalo-Hernández, C. O., Baligar, V. C., & He, Z. L. (2017). Heavy metal accumulation in leaves and beans of cacao (Theobroma cacao L.) in major cacao growing regions in Peru. *Science of the Total Environment, 605*–606(2017), 792–800.

Assa, A., Noor, A., Yunus, M. R., Misnawi, & Djide, M. N. (2018). Heavy metal concentrations in cocoa beans (Theobroma cacao L.) originating from East Luwu, South Sulawesi, Indonesia. *Journal of Physics: Conference Series*, *979*(1).

Barbin, D. F., Maciel, L. F., Bazoni, C. H. V, Ribeiro, M. D. S., Carvalho, R. D. S., Bispo, E. D. S., Miranda, M. D., Hirooka, E. Y. (2018). Classification and compositional characterization of different varieties of cocoa beans by near infrared spectroscopy and multivariate statistical analyses. *Journal of Food Science and Technology*, 1–10.

Belo, R. F. C., Figueiredo, J. P., Nunes, C. M., Pissinatti, R., Souza, S. V. C. de, & Junqueira, R. G. (2017). Accelerated solvent extraction method for the quantification of polycyclic aromatic hydrocarbons in cocoa beans by gas chromatography–mass spectrometry. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, *1053*, 87–100.

Belščak, A., Komes, D., Horžić, D., Ganić, K. K., & Karlović, D. (2009). Comparative study of commercially available cocoa products in terms of their bioactive composition. *Food Research International*, *42*(5–6), 707–716.

Berrueta, L. A., Alonso, R. M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A*, *1158*(1–2), 196–214.

Beulens, A. J. M., Broens, D. F., Folstar, P., & Hofstede, G. J. (2005). Food safety and transparency in food chains and networks. Relationships and challenges. *Food Control*, *16*(6 SPEC. ISS.), 481–486.

Bolliger, S., Zeng, Y., & Windhab, E. J. (1999). In-line Measurement of Tempered Cocoa Butter and Chocolate, *JAOCS, Journal of the American Oils Chemist' Society, 76*(6), 659–667.

Bonvehí, J. S. (2005). Investigation of aromatic compounds in roasted cocoa powder. *European Food Research and Technology*, *221*(1–2), 19–29.

Bratinova, S., Karasek, L., Buttinger, G., & Wenzl, T. (2015). Report on the 16th Interlaboratory comparison organnnsed by the European Union Reference Laboraty for Polycyclic Aromatic Hydrocarbons, EUR 27558 15. EU. https://doi.org/10.2787/279750.

Brera, C., Grossi, S., & Miraglia, M. (2005). Interlaboratory study for

ochratoxin A determination in cocoa powder samples. *Journal of Liquid Chromatography and Related Technologies*, *28*(1), 35–61.

Bro, R. (1997). PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems, 38*, 149–171.

Cádiz-Gurrea, M. L., Lozano-Sanchez, J., Contreras-Gámez, M., Legeai-Mallet, L., Fernández-Arroyo, S., & Segura-Carretero, A. (2014). Isolation, comprehensive characterization and antioxidant activities of Theobroma cacao extract. *Journal of Functional Foods*, *10*, 485–498.

Cambrai, A., Marcic, C., Morville, S., Houer, P. S., Bindler, F., & Marchioni, E. (2010). Differentiation of chocolates according to the cocoa's geographical origin using chemometrics. *Journal of Agricultural and Food Chemistry*, *58*(3), 1478–1483.

CAOBISCO-ECA-FCC (2015). Cocoa beans: chocolate and cocoa industry quality requirements. (End, M. J. and Dand, R., eds.). http://www.cocoaquality.eu/

Caporaso, N., Whitworth, M. B., Fowler, M. S., & Fisk, I. D. (2018). Hyperspectral imaging for non-destructive prediction of fermentation index, polyphenol content and antioxidant activity in single cocoa beans. *Food Chemistry*, *258*(March), 343–351.

CBI. (2016). *CBI Trade Statistics: Cocoa in Europe*. Netherlands. Retrieved from https://www.cbi.eu/sites/default/files/market_information/researches/trade-statistics-europe-cocoa-2016.pdf

Chavez, E., He, Z. L., Stoffella, P. J., Mylavarapu, R. S., Li, Y. C., Moyano, B., & Baligar, V. C. (2015). Concentration of cadmium in cacao beans and its relationship with soil cadmium in southern Ecuador. *Science of the Total Environment*, *533*, 205–214.

Chavez, E., He, Z. L., Stoffella, P. J., Mylavarapu, R. S., Li, Y. C., & Baligar, V. C. (2016). Chemical speciation of cadmium: An approach to

evaluate plant-available cadmium in Ecuadorian soils under cacao production. *Chemosphere, 150*, 57–62.

Chetschik, I., Kneubühl, M., Chatelain, K., Schlüter, A., Bernath, K., & Hühn, T. (2018). Investigations on the Aroma of Cocoa Pulp (Theobroma cacao L.) and Its Influence on the Odor of Fermented Cocoa Beans. *Journal of Agricultural and Food Chemistry, 66*(10), 2467–2472.

Codex Alimentarius. (2014). Codex Alimentarius Cocoa- Cocoa liquor.

CODEX STAN 228-2001. (2001). GENERAL METHODS OF ANALYSIS FOR CONTAMINANTS CODEX STAN 228-2001.

Cordella, M., Torri, C., Adamiano, A., Fabbri, D., Barontini, F., & Cozzani, V. (2012). Bio-oils from biomass slow pyrolysis: A chemical and toxicological screening. *Journal of Hazardous Materials, 231–232*, 26–35.

Cortés, V., Blasco, J., Aleixos, N., Cubero, S., & Talens, P. (2019). Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. *Trends in Food Science & Technology, 85*(January), 138–148.

Counet, C., Ouwerx, C., Rosoux, D., & Collin, S. (2004). Relationship between procyanidin and flavor contents of cocoa liquors from different origins. Journal of agricultural and food chemistry, 52(20), 6243-6249.

Crafack, M., Keul, H., Eskildsen, C. E., Petersen, M. A., Saerens, S., Blennow, A., Heimdal, H., Skovnan, L. M., Siewers, J. H,, Nielsen, D. S. (2014). Impact of starter cultures and fermentation techniques on the volatile aroma and sensory profile of chocolate. *Food Research International, 63*, 306–316.

Crouzillat, D., Bellanger, L., Rigoreau, M., Bucheli, P., & Pétiard, V. (2000). Genetic structure, characterisation and selection of Nacional cocoa compared to other genetic groups. In international workshop on new technologies and cocoa breeding (p. 30).

Introduction

Cubero-Leon, E., Bouten, K., Senyuva, H., & Stroka, J. (2017). Determination of ochratoxin a in black and white pepper, nutmeg, spice mix, cocoa, and drinking chocolate by high-performance liquid chromatography coupled with fluorescence detection: Collaborative study. *Journal of AOAC International*, *100*(5), 1458–1468.

D'Souza, R. N., Grimbs, S., Behrends, B., Bernaert, H., Ullrich, M. S., & Kuhnert, N. (2017). Origin-based polyphenolic fingerprinting of Theobroma cacao in unfermented and fermented beans. *Food Research International*, *99*(May), 550–559.

Di Mattia, C., Martuscelli, M., Sacchetti, G., Beheydt, B., Mastrocola, D., & Pittia, P. (2014). Effect of different conching processes on procyanidin content and antioxidant properties of chocolate. *Food Research International*, *63*, 367–372.

Dickens, B., & Dickens, S. (1999). Estimation of concentration and bonding environment of water dissolved in common solvents using near infrared absorptivity. *Journal of Research-National Institute of Standards and Technology*, *104*(2), 173–183. Retrieved from http://nvlpubs.nist.gov/nistpubs/jres/104/2/html/dickens-n/dickens.htm

Diem, P., Walle, D. Van De, Clercq, N. De, Winne, A. De, Kadow, D., Lieberei, R., Messens, K., Nhan, D., Dwelttinck, K., Durme, J. Van (2015). Assessing cocoa aroma quality by multiple analytical approaches. *FRIN*, *77*, 657–669.

Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A., & Smith, F. (1956). Colorimetric Method for Determination of Sugars and Related Substances. *Analytical Chemistry*, 28(3), 350–356.

Elwers, S., Zambrano, A., Rohsius, C., & Lieberei, R. (2009). Differences between the content of phenolic compounds in Criollo, Forastero and Trinitario cocoa seed (Theobroma cacao L.). *European Food Research and Technology*, *229*(6), 937–948.

62

EU. (2000). EU Directive 2000/36/EC of the European Parliament and of the Council of 23 June 2000 relating to cocoa and chocolate products intended for human consumption. *Official Journal of the European Communities*, *L197*, 19–25.

European Commission. (2006). Commission Regulation (EC) No 1881/2006 of 19 December 2006 setting maximum levels for certain contaminants in foodstuffs. *Official Journal,* (1881/2006), 5–24.

European Commission. (2010). Commission Regulation (EU) No 105/ 2010 of 5 February 2010 amending Regulation (EC) No 1881/ 2006 setting maximum levels for certain contaminants in food- stuffs as regards ochratoxin A. Off J Eur Union 35:7–8

European Commission. (2011). Commission regulation (EU) No 835/2011 of 19 August 2011 amending Regulation (EC) No 1881/2006 as regards maximum levels for polycyclic aromatic hydrocarbons in foodstuffs, Official Journal of the European Union (2011).

Fayeulle, N., Vallverdu-Queralt, A., Meudec, E., Hue, C., Boulanger, R., Cheynier, V., & Sommerer, N. (2018). Characterization of new flavan-3-ol derivatives in fermented cocoa beans. *Food Chemistry*, *259*(March), 207–212.

FCC. (2018). Services - Rules | The Federation of Cocoa Commerce. Retrieved June 20, 2018, from http://www.cocoafederation.com/services/rules

Forsyth, W. G. C., & Quesnel, V. C. (1957). Cacao polyphenolic substances. 4. The anthocyanin pigments. *Biochemical Journal, 65*(1), 177–179.

Franco, R., Oñatibia-Astibia, A., & Martínez-Pinilla, E. (2013). Health benefits of methylxanthines in cacao and chocolate. *Nutrients*, *5*(10), 4159–4173.

Gabbay Alves, T. V., Silva da Costa, R., Aliakbarian, B., Casazza, A. A., Perego, P., Pinheiro Arruda, M. S., Carrerá, J.O., Converti, A., Ribeiro, R. M.

Introduction

(2017). Bioactive compounds and antioxidant potential for polyphenol-rich cocoa extract obtained by agroindustrial residue. *Natural Product Research*, *6419*(November), 1–4.

García-Alamilla, P., Salgado-Cervantes, M. A., Barel, M., Berthomieu, G., Rodríguez-Jimenes, G. C., & García-Alvarado, M. A. (2007). Moisture, acidity and temperature evolution during cacao drying. *Journal of Food Engineering*, *79*(4), 1159–1165.

Gianfredi, V., Salvatori, T., Nucci, D., Villarini, M., & Moretti, M. (2018). Can chocolate consumption reduce cardio-cerebrovascular risk? A systematic review and meta-analysis. *Nutrition*, *46*, 103–114.

Goodacre, R., & Anklam, E. (2001). Fourier Transform infrared Spectroscopy and Chemometrics as a Tool for the Rapid Detection of Other Vegetable Fats Mixed in Cocoa Butter, *78*(10).

Guehi, S. T., Dingkuhn, M., Cros, E., Fourny, G., Moulin, G., & Vidal, A. C. (2008). Impact of cocoa processing technologies in free fatty acids formation in stored raw cocoa beans, *3*(March), 174–179.

Hashimoto, J. C., Lima, J. C., Celeghini, R. M. S., Nogueira, A. B., Efraim, P., Poppi, R. J., & Pallone, J. A. L. (2018). Quality Control of Commercial Cocoa Beans (Theobroma cacao L.) by Near-infrared Spectroscopy. *Food Analytical Methods*, *11*(5), 1510–1517.

Hinneh, M., Semanhyia, E., Van de Walle, D., De Winne, A., Tzompa-Sosa, D. A., Scalone, G. L. L., De Meuleanaer, B., Messens, K., Van Durme, J., Afoakwa, E. O., De Cooman, L., Dewettinck, K. (2018). Assessing the influence of pod storage on sugar and free amino acid profiles and the implications on some Maillard reaction related flavor volatiles in Forastero cocoa beans. *Food Research International, 111*(January), 607–620.

Huang, X. Y., Teye, E., Sam-Amoah, L. K., Han, F. K., Yao, L. Y., & Tchabo, W. (2014). Rapid measurement of total polyphenols content in cocoa beans by data fusion of NIR spectroscopy and electronic tongue. *Analytical Methods*, *6*(14), 5008–5015.

Hue, C., Gunata, Z., Bergounhou, A., Assemat, S., Boulanger, R., Sauvage, F. X., & Davrieux, F. (2014). Near infrared spectroscopy as a new tool to determine cocoa fermentation levels through ammonia nitrogen quantification. *Food Chemistry*, *148*, 240–245.

Hue, C., Gunata, Z., Breysse, A., Davrieux, F., Boulanger, R., & Sauvage, F. X. (2016). Impact of fermentation on nitrogenous compounds of cocoa beans (Theobroma cacao L.) from various origins. *Food Chemistry*, *192*, 958–964.

Humston, E. M., Knowles, J. D., McShea, A., & Synovec, R. E. (2010). Quantitative assessment of moisture damage for cacao bean quality using two-dimensional gas chromatography combined with time-of-flight mass spectrometry and chemometrics. *Journal of Chromatography A, 1217*(12), 1963–1970.

ICCO. (2012). Physical and chemical information on cocoa beans, butter, mass and powder. Retrieved August 13, 2018, from https://www.icco.org/faq/61-physical-and-chemical-information-on-cocoa/106-physical-and-chemical-information-on-cocoa-beans-butter-mass-and-powder.html

ICCO. (2018). How is the quality of cocoa checked - by hand, by machine? Retrieved June 20, 2018, from https://www.icco.org/faq/59-fermentation-a-drying/108-how-is-the-quality-of-cocoa-checked-by-hand-by-machine.html.

Introduction

ICCO. (2019). Leading countries of cocoa bean processing worldwide, 2018/2019 | Statista. Retrieved November 4, 2019, from https://www.statista.com/statistics/238242/leading-countries-of-global-cocoa-bean-processing/

Ioannone, F., Di Mattia, C. D., De Gregorio, M., Sergi, M., Serafini, M., & Sacchetti, G. (2015). Flavanols, proanthocyanidins and antioxidant activity changes during cocoa (Theobroma cacao L.) roasting as affected by temperature and time of processing. Food Chemistry, 174, 256–262.

IOCCC. (1996). *International Office of Chocolate*. Determination of free fatty acids (FFA) content of cocoa fat as a measure of cocoa nib acidity. Analytical Method N° 42-1993.

Ishaq, S., & Jafri, L. (2017). Biomedical Importance of Cocoa (Theobroma cacao): Significance and Potential for the Maintenance of Human Health. *Matrix Science Pharma*, *1*(1), 1–5.

Jackson, E., Farrington, D. S., & Henderson, K. (1986). The analysis of agricultural materials: a manual of the analytical methods used by the Agricultural Development and Advisory Service. *The Analysis of Agricultural Materials: A Manual of the Analytical Methods Used by the Agricultural Development and Advisory Service.*, (No. 427 (Ed. 3)).

Jahurul, M. H. A., Soon, Y., Shaarani Sharifudin, M., Hasmadi, M., Mansoor, A. H., Zaidul, I. S. M., Lee, J. S., Ali, M. E., Ghafoor, K., Zzaman, W., Jinap, S. (2018). Bambangan (Mangifera pajang) kernel fat: A potential new source of cocoa butter alternative. *International Journal of Food Science and Technology*, 1689–1697.

Jinap, S., Thien, J., & Yap, T. N. (1994). Effect of drying on acidity and volatile fatty-acids content of cocoa beans. *Journal of the Science of Food and Agriculture*, *65*(1), 67–75.

Kongor, J. E., Hinneh, M., de Walle, D. Van, Afoakwa, E. O., Boeckx, P., & Dewettinck, K. (2016). Factors influencing quality variation in cocoa

(Theobroma cacao) bean flavour profile - A review. *Food Research International*, *82*, 44–52.

Kostic, M. J. (1997). Cocoa Alkalization. *The Manufacturing Confectioner*, (June), 128–130.

Krähmer, A., Engel, A., Kadow, D., Ali, N., Umaharan, P., Kroh, L. W., & Schulz, H. (2015). Fast and neat – Determination of biochemical quality parameters in cocoa using near infrared spectroscopy. *Food Chemistry, 181*, 152–159.

Krähmer, Gudi, G., Weiher, N., Gierus, M., Schütze, W., & Schulz, H. (2013). Characterization and quantification of secondary metabolite profiles in leaves of red and white clover species by NIR and ATR-IR spectroscopy. *Vibrational Spectroscopy*, *68*, 96–103.

Kruszewski, B., Obiedziński, M. W., & Kowalska, J. (2018). Nickel, cadmium and lead levels in raw cocoa and processed chocolate mass materials from three different manufacturers. *Journal of Food Composition and Analysis*, *66*(March 2017), 127–135.

Kubíčková, A., Kubíček, V., & Coufal, P. (2011). UV-VIS detection of amino acids in liquid chromatography: Online post-column solid-state derivatization with Cu(II) ions. *Journal of Separation Science*, *34*(22), 3131–3135.

Kucha, C. T., Liu, L., & Ngadi, M. O. (2018). Non-destructive spectroscopic techniques and multivariate analysis for assessment of fat quality in pork and pork products: A review. *Sensors (Switzerland)*, *18*(2).

Kumari, N., Grimbs, A., D'Souza, R. N., Verma, S. K., Corno, M., Kuhnert, N., & Ullrich, M. S. (2018). Origin and varietal based proteomic and peptidomic fingerprinting of Theobroma cacao in non-fermented and fermented cocoa beans. *Food Research International*, *111*(April), 137–147.

Introduction

Kutsanedzie, F. Y. H., Chen, Q., Hassan, M. M., Yang, M., Sun, H., & Rahman, M. H. (2018). Near infrared system coupled chemometric algorithms for enumeration of total fungi count in cocoa beans neat solution. *Food Chemistry*, *240*(May 2017), 231–238.

Lacueva, C. A., Monagas, M., Khan, N., Izquterdo-Pulido, M., Urpi-Sarda, M., Permanyer, J., & Lamuela-Raventós, R. M. (2008). Flavanol and flavonol contents of cocoa powder products: Influence of the manufacturing process. *Journal of Agricultural and Food Chemistry*, *56*(9), 3111–3117.

Langer, S., Marshall, L. J., Day, A. J., & Morgan, M. R. A. (2011). Flavanols and methylxanthines in commercially available dark chocolate: A study of the correlation with nonfat cocoa solids. *Journal of Agricultural and Food Chemistry*, *59*(15), 8435–8441.

Lee, S. L., & Djauhari, M. A. (2013). Quality control in cocoa powder production process: A robust MSPC approach. *Jurnal Teknologi (Sciences and Engineering)*, *63*(2), 41–44.

Levasseur-garcia, C. (2018). Updated Overview of Infrared Spectroscopy Methods for Detecting Mycotoxins on Cereals (Corn, Wheat, and Barley). Review. *Toxins, 10,* 38.

Li, Y., Feng, Y., Zhu, S., Luo, C., Ma, J., & Zhong, F. (2012). The effect of alkalization on the bioactive and flavor related components in commercial cocoa powder. *Journal of Food Composition and Analysis*, *25*(1), 17–23.

Li, Y., Zhu, S., Feng, Y., Xu, F., Ma, J., & Zhong, F. (2014). Influence of alkalization treatment on the color quality and the total phenolic and anthocyanin contents in cocoa powder. *Food Science and Biotechnology*, *23*(1), 59–63.

Li, Yao, W., Liu, Q., Xu, J., Bao, B., Shan, M., Cao, Y., Ding, A., Cheng, F., Zhang, L. (2017). Application of UHPLC-ESI-Q-TOF-MS to identify multiple constituents in processed products of the herbal medicine ligustri lucidi fructus. *Molecules*, *22*(5).

Liu, M., Liu, J., He, C., Song, H., Liu, Y., Zhang, Y., Wang, E., Guo, J., Yang, H., Su, X. (2017). Characterization and comparison of key aroma-active compounds of cocoa liquors from five different areas. *International Journal of Food Properties*, *20*(10), 2396–2408.

Lo Dico, G. M. L., Galvano, F., Dugo, G., D'ascenzi, C., Macaluso, A., Vella, A., & Ferrantelli, V. (2018). Toxic metal levels in cocoa powder and chocolate by ICP-MS method after microwave-assisted digestion. *Food Chemistry, 245*, 1163–1168.

Loullis, A., & Pinakoulaki, E. (2018). Carob as cocoa substitute: a review on composition, health benefits and food applications. *European Food Research and Technology*, *244*(6), 959–977.

Luna, F., Crouzillat, D., Cirou, L., & Bucheli, P. (2002) Chemical composition and flavor of Ecuadorian cocoa liquor. Journal of Agricultural and Food Chemistry, 50 (12), 3527-3532.

Machonis, P., Jones, M., Schaneberg, B., Kwik-Uribe, C., & Dowell, D. (2014). Method for the determination of catechin and epicatechin enantiomers in cocoa-based ingredients and products by high-performance liquid chromatography: First action 2013.04. *Journal of AOAC International*, *97*(2), 506–509.

Magagna, F., Guglielmetti, A., Liberto, E., Reichenbach, S. E., Allegrucci, E., Gobino, G., Bicchi, C., Cordero, C. (2017). Comprehensive Chemical Fingerprinting of High-Quality Cocoa at Early Stages of Processing: Effectiveness of Combined Untargeted and Targeted Approaches for Classification and Discrimination. *Journal of Agricultural and Food Chemistry*, *65*(30), 6329–6341.

Manning, L. (2016). Food fraud: policy and food chain. *Current Opinion in Food Science*, *10*(2), 16–21.

Manzano, P., Hernández, J., Quijano-Avilés, M., Barragán, A., Chóez-Guaranda, I., Viteri, R., & Valle, O. (2017). Polyphenols extracted from

Introduction

Theobroma cacao waste and its utility as antioxidant. *Emirates Journal of Food and Agriculture, 29*(1), 45–50.

Martens, H., Nielsen, J. P., & Engelsen, S. B. (2003). Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures. *Analytical Chemistry, 75*(3), 394–404.

Martín, M. Á., & Ramos, S. (2017). Health beneficial effects of cocoa phenolic compounds: a mini-review. *Current Opinion in Food Science, 14,* 20–25.

Mendes & Lima, E. L. (2007). Perfil Agroindustrial do Processamento de Amêndoas de Cacau em Pequena Escala no Estado do Pará, Belém. *SEMBRAE/PA.*

Miller, K. B., Hurst, W. J., Payne, M. J., Stuart, D. A., Apgar, J., Sweigart, D. S., & Ou, B. (2008). Impact of alkalization on the antioxidant and flavanol content of commercial cocoa powders. *Journal of Agricultural and Food Chemistry, 56*(18), 8527–8533.

Miller, K. B., Stuart, D. A., Smith, N. L., Lee, C. Y., Mchale, N. L., Flanagan, J. A., Boxing, O. U., Hurst, W. J. (2006). Antioxidant activity and polyphenol and procyanidin contents of selected commercially available cocoa-containing and chocolate products in the United States. *Journal of Agricultural and Food Chemistry, 54*(11), 4062–4068.

Moros, J., Iñón, F. A., Garrigues, S., & de la Guardia, M. (2007). Near-infrared diffuse reflectance spectroscopy and neural networks for measuring nutritional parameters in chocolate samples. *Analytica Chimica Acta, 584*(1), 215–222.

Nielsen, D. S., Snitkjaer, P., & Van Den Berg, F. (2008). Investigating the fermentation of cocoa by correlating Denaturing Gradient Gel Electrophoresis profiles and Near Infrared spectra. *International Journal of Food Microbiology, 125,* 133–140.

Niemenak, N., Rohsius, C., Elwers, S., Omokolo Ndoumou, D., & Lieberei, R. (2006). Comparative study of different cocoa (Theobroma cacao L.) clones in terms of their phenolics and anthocyanins contents. *Journal of Food Composition and Analysis*, *19*(6–7), 612–619.

Okiyama, D. C. G., Navarro, S. L. B., & Rodrigues, C. E. C. (2017). Trends in Food Science & Technology Cocoa shell and its compounds : Applications in the food industry. *Trends in Food Science & Technology*, *63*, 103–112.

Oliviero, T., Capuano, E., Ca, B., & Fogliano, V. (2009). Influence of Roasting on the Antioxidant Activity and HMF Formation of a Cocoa Bean Model Systems Influence of Roasting on the Antioxidant Activity and HMF Formation of a Cocoa Bean Model Systems. *Society*, 147–152.

Oracz, J., Nebesny, E., & Żyżelewicz, D. (2015). Changes in the flavan-3-ols, anthocyanins, and flavanols composition of cocoa beans of different Theobroma cacao L. groups affected by roasting conditions. *European Food Research and Technology*, *241*(5), 663–681.

Osborne, B. G. (2000). Near-Infrared Spectroscopy in Food Analysis. *In Encyclopedia of Analytical Chemistry* (pp. 1–14). 2006 John Wiley & Sons, Ltd.

Payne, M. J., Hurst, W. J., Miller, K. B., Rank, C., & Stuart, D. A. (2010). Impact of fermentation, drying, roasting, and dutch processing on epicatechin and catechin content of cacao beans and cocoa ingredients. *Journal of Agricultural and Food Chemistry*, *58*(19), 10518–10527.

Pedan, V., Fischer, N., Bernath, K., Hühn, T., & Rohn, S. (2017). Determination of oligomeric proanthocyanidins and their antioxidant capacity from different chocolate manufacturing stages using the NP-HPLC-online-DPPH methodology. *Food Chemistry*, *214*, 523–532.

Pedan, V., Fischer, N., & Rohn, S. (2016). Extraction of cocoa proanthocyanidins and their fractionation by sequential centrifugal partition

chromatography and gel permeation chromatography. *Analytical and Bioanalytical Chemistry, 408*(21), 5905–5914.

Pérez, É., Lerma, M. J., Fuentes, A., Palomares, C., & Barat, J. M. (2016). Control of undeclared flavoring of cocoa powders by the determination of vanillin and ethyl vanillin by HPLC. *Food Control, 67*, 171–176.

Permanyer, J. J., & Perez, M. L. (1989). Compositional analysis of powdered coca products by Near infrared reflectance spectroscopy. Journal of *Food Science, 54*(3), 768–769.

Popp, J., Rösch, P., Ramos, L. A., Salinas, W., Ciobotă, V., Aponte, P. M., Vargas., P., Kampe, B. (2016). Distinction of Ecuadorian varieties of fermented cocoa beans using Raman spectroscopy. *Food Chemistry, 211*, 274–280.

Quelal-Vásconez, M. A., Lerma-García, M. J., Pérez-Esteve, É., Arnau-Bonachera, A., Barat, J. M., & Talens, P. (2019). Fast detection of cocoa shell in cocoa powders by near infrared spectroscopy and multivariate analysis. *Food Control, 99*(November 2018), 68–72.

Quelal-Vásconez, M. A., Pérez-Esteve, É., Arnau-Bonachera, A., Barat, J. M., & Talens, P. (2018). Rapid fraud detection of cocoa powder with carob flour using near infrared spectroscopy. *Food Control, 92*, 183–189.

Quelal-Vásconez, M. A., Lerma-García, M. J., Pérez-Esteve, É., Arnau-Bonachera, A., Barat, J. M., & Talens, P. (2020). Changes in methylxanthines and flavanols during cocoa powder processing and their quantification by near-infrared spectroscopy. *LWT, 117*(December 2018), 108598.

Quiroz-Reyes, C. N., & Fogliano, V. (2018). Design cocoa processing towards healthy cocoa products: The role of phenolics and melanoidins. *Journal of Functional Foods, 45*(January), 480–490.

Ramalingam, C., Jain, H., Vatsa, K., Akhtar, N., Mitra, B., Vishnudas, D., Yadav, S., Garg, K., Prakash, A., Rai, A. (2009). *Detection and Biochemical Characterization of Microorganisms in Milk and Cocoa powder*

samples by FTIR and subsequent production of Bacteriocin from *Lactobacillus*. *International Journal of Drug Development and Research* (Vol. 5). International Journal of Drug Development & Research. Retrieved from http://www.ijddr.in/drug-development/detection-and-biochemical-characterization-of-microorganisms-in-milk-and-cocoa-powder-samples-by-ftir-and-subsequent-production-of-bacteriocin-from-lactobacillus.php?aid=6647

Reddy, S. Y., & Prabhakar, J. V. (1994). Cocoa butter extenders from Kokum (*Garcinia indica*) and Phulwara (*Madhuca butyracea*) butter. *Journal of the American Oil Chemists' Society*, *71*(2), 217–219.

Risner, C. H. (2008). Simultaneous Determination of Theobromine, (+)-Catechin, Caffeine, and (-)-Epicatechin in Standard Reference Material Baking Chocolate 2384, Cocoa, Cocoa Beans, and Cocoa Butter. *Journal of Chromatographic Science*, *46*(10), 892–899.

Rodrigues, J. F., Condino, J. P. F., Pinheiro, A. C. M., & Nunes, C. A. (2016). Temporal dominance of sensations of chocolate bars with different cocoa contents: Multivariate approaches to assess TDS profiles. *Food Quality and Preference, 47,* 91–96.

Rodriguez-Campos, J., Escalona-Buendía, H. B., Contreras-Ramos, S. M., Orozco-Avila, I., Jaramillo-Flores, E., & Lugo-Cervantes, E. (2012). Effect of fermentation time and drying temperature on volatile compounds in cocoa. *Food Chemistry*, *132*(1), 277–288.

Rodríguez-Carrasco, Y., Gaspari, A., Graziani, G., Sandini, A., & Ritieni, A. (2018). Fast analysis of polyphenols and alkaloids in cocoa-based products by ultra-high performance liquid chromatography and Orbitrap high resolution mass spectrometry (UHPLC-Q-Orbitrap-MS/MS). *Food Research International*, *111*(May), 229–236.

Rohsius, C., Matissek, R., & Lieberei, R. (2006). Free amino acid amounts in raw cocoas from different origins. *European Food Research and*

Technology, *222*(3–4), 432–438.

Salem, E. M., & Ohaad Fahad, A. A. (2012). Substituting of Cacao by Carob Pod Powder In Milk Chocolate Manufacturing. *Australian Journal of Basic and Applied Sciences*, *6*(3), 572–578.

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, *36*(8), 1627–1639.

Schwan, R. F. (1998). Cocoa fermentations conducted with a defined microbial cocktail inoculum. *Applied and Environmental Microbiology*, *64*(4), 1477–1483.

Servent, A., Boulanger, R., Davrieux, F., Pinot, M., Tardan, E., Forestier-chiron, N., & Hue, C. (2018). Assessment of cocoa (Theobroma cacao L .) butter content and composition throughout fermentations. *Food Research International*, *107*(March), 675–682.

Sess-Tchotch, D.-A., Kedjebo, K. B. D., Faulet, B. M., Fontana-Tachon, A., Alter, P., Durand, N., Grabulos, J., Montet, D., Guehi, T. S. (2018). Analytical Method Validation and Rapid Determination of Polycyclic Aromatic Hydrocarbons (PAHs) in Cocoa Butter Using HPLC-FLD. *Food Analytical Methods*.

Shavez, M., Ahmad, S., Jan, K., & Bashir, K. (2017). Trends in Food Science & Technology Status , supply chain and processing of cocoa - A review. *Trends in Food Science & Technology*, *66*, 108–116.

Soto, J., Granda, G., Prieto, F., Ipanaque, W., & Machacuay, J. (2015). Cocoa bean quality assessment by using hyperspectral images and fuzzy logic techniques. *In Twelfth International Conference on Quality Control by Artificial Vision* (Vol. 9534, p. 95340M).

Stark, T., Lang, R., Keller, D., Hensel, A., & Hofmann, T. (2008). Absorption of N-phenylpropenoyl-L-amino acids in healthy humans by oral administration of cocoa (Theobroma cacao). *Molecular Nutrition and Food*

Research, *52*(10), 1201–1214.

Su, W.-H., & Sun, D.-W. (2017). Fourier Transform Infrared and Raman and Hyperspectral Imaging Techniques for Quality Determinations of Powdery Foods: A Review. *Comprehensive Reviews in Food Science and Food Safety*, *17*.

Suazo, Y., Davidov-Pardo, G., & Arozarena, I. (2014). Effect of Fermentation and Roasting on the Phenolic Concentration and Antioxidant Activity of Cocoa from Nicaragua. *Journal of Food Quality*, *37*(1), 50–56.

Sunoj, S., Igathinathane, C., & Visvanathan, R. (2016). Nondestructive determination of cocoa bean quality using FT-NIR spectroscopy. *Computers and Electronics in Agriculture*, *124*, 234–242.

Talbot, C. P. J., Mensink, R. P., Smolders, L., Bakeroot, V., & Plat, J. (2018). Theobromine does not affect fasting and postprandial HDL cholesterol efflux capacity , while it decreases fasting miR-92a levels in humans, 1–25.

Tan, J., & Kerr, W. L. (2018). Determining degree of roasting in cocoa beans by artificial neural network (ANN)-based electronic nose system and gas chromatography/mass spectrometry (GC/MS). *Journal of the Science of Food and Agriculture*, (January).

Teye, E., & Huang, X. (2015). Novel Prediction of Total Fat Content in Cocoa Beans by FT-NIR Spectroscopy Based on Effective Spectral Selection Multivariate Regression. *Food Analytical Methods*, *8*(4), 945–953.

Teye, E., Huang, X., Dai, H., & Chen, Q. (2013). Spectrochimica Acta Part A : Molecular and Biomolecular Spectroscopy Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *114*, 183–189.

Teye, E., Huang, X., Han, F., & Botchway, F. (2014). Discrimination of Cocoa Beans According to Geographical Origin by Electronic Tongue and

Introduction

Multivariate Algorithms. *Food Analytical Methods*, *7*(2), 360–365.

Teye, E., Huang, X., Lei, W., & Dai, H. (2014). Feasibility study on the use of Fourier transform near-infrared spectroscopy together with chemometrics to discriminate and quantify adulteration in cocoa beans. *FRIN*, *55*, 288–293.

Teye, E., Huang, X., Sam-Amoah, L. K., Takrama, J., Boison, D., Botchway, F., & Kumi, F. (2015). Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis. *Food Chemistry*, *176*, 403–410.

Teye, E., Huang, X., Takrama, J., & Haiyang, G. (2014). Integrating NIR spectroscopy and electronic tongue together with chemometric analysis for accurate classification of cocoa bean varieties. *Journal of Food Process Engineering*, *37*(6), 560–566.

Teye, E., Uhomoibhi, J., & Wang, H. (2016). Nondestructive Authentication of Cocoa Bean Cultivars by FT-NIR Spectroscopy and Multivariate Techniques. *Focus on Science*, *2*(3).

Toro-Uribe, S., Montero, L., López-Giraldo, L., Ibáñez, E., & Herrero, M. (2018). Characterization of secondary metabolites from green cocoa beans using focusing-modulated comprehensive two-dimensional liquid chromatography coupled to tandem mass spectrometry. *Analytica Chimica Acta*.

Torres-Moreno, M., Tarrega, A., & Blanch, C. (2014). *Characterization of Volatile Compounds in Dark Chocolates by HS-SPME and GC-MS. Flavour Science*. Elsevier Inc.

Torres-Moreno, M., Torrescasana, E., Salas-Salvadó, J., & Blanch, C. (2015). Nutritional composition and fatty acids profile in cocoa beans and chocolates with different geographical origin and processing conditions. *Food Chemistry*, *166*, 125–132.

Trafialek, J., & Kolanowski, W. (2017). Article information: Implementation and functioning of HACCP principles in certified and non-

certified food businesses: A preliminary study. *British Food Journal, 119(*4), 710–728.

Trilčová, A., Čopíková, J., A Coimbra, M., Barros, A., Egert, L., Synytsya, A., & Křístková, H. (2004). Application of NIR analysis to verify cocoa powder authenticity. *Czech Journal of Food Sciences, 22*(SI-Chem. Reactions in Foods V), S329–S332.

Turcotte, A., Scott, P. M., & Tague, B. (2013). Analysis of cocoa products for ochratoxin A and aflatoxins, 193–201.

Valdez, L.F., & Gutiérrez, J.M. (2016). Chocolate classification by an electronic nose with pressure controlled generated stimulation. *Sensors (Switzerland), 16*(10).

Van Durme, J., Ingels, I., & De Winne, A. (2016). Inline roasting hyphenated with gas chromatography-mass spectrometry as an innovative approach for assessment of cocoa fermentation quality and aroma formation potential. *Food Chemistry*, *205*, 66–72.

Van Ruth, S. M., Luning, P. A., Silvis, I. C. J., Yang, Y., & Huisman, W. (2018). Differences in fraud vulnerability in various food supply chains and their tiers. *Food Control*, *84*, 375–381.

Vázquez-Ovando, A., Chacón-Martínez, L., Betancur-Ancona, D., Escalona-Buendía, H., & Salvador-Figueroa, M. (2015). Sensory descriptors of cocoa beans from cultivated trees of Soconusco, Chiapas, Mexico. *Food Science and Technology (Campinas)*, *35*(2), 285–290.

Veselá, A., Barros, A. S., Synytsya, A., Delgadillo, I., Čopíková, J., & Coimbra, M. A. (2007). Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder. *Analytica Chimica Acta*, *601*(1), 77–86.

Voigt, J., Janek, K., Textoris-Taube, K., Niewienda, A., & Wöstemeyer, J. (2016). Partial purification and characterisation of the peptide precursors of the cocoa-specific aroma components. *Food Chemistry*, *192*, 706–713.

Introduction

Voigt, J., Textoris-Taube, K., & Wöstemeyer, J. (2018). pH-Dependency of the proteolytic formation of cocoa- and nutty-specific aroma precursors. *Food Chemistry, 255*(February), 209–215.

Weiller, S., Tanabe, T., & Oyama, Y. (2018). Terahertz Non-Contact Monitoring of Cocoa Butter in Chocolate *. *World Journal of Engineering and Technology, 6*, 268–274.

Whitacre, E., Oliver, J., Van Den Broek, R., Engelsen, P., Kremers, B., Van Der Horst, B., Steward, M., Jansen- Beuwink, A. (2003). Predictive Analysis of Cocoa Procyanidins Using Near-Infrared Spectroscopy Techniques. *Food Chemistry and Toxicology Predictive, 68*(9), 2618–2622.

Wold, S., Antti, H., Lindgren, F., & Öhman, J. (1998). Orthogonal signal correction of near-infrared spectra. *Chemometrics and Intelligent Systems, 44*, 175–185. Retrieved from http://www.sciencedirect.com/science/article/pii/S0169743998001099

Wollgast, J., & Anklam, E. (2000). Review on polyphenols in Theobroma cacao : changes in composition during the manufacture of chocolate and methodology for identication and quantification. *Food Research International, 33*(July).

Yang, W. li, Hu, M. hua, Chen, S. wei, Wang, Q., Zhu, S., Dai, J., & Li, X. zhong. (2015). Identification of Adulterated Cocoa Powder Using Chromatographic Fingerprints of Polysaccharides Coupled with Principal Component Analysis. *Food Analytical Methods, 8*(9), 2360–2367.

Yılmaz, C., & Gökmen, V. (2018). Determination of tryptophan derivatives in kynurenine pathway in fermented foods using liquid chromatography tandem mass spectrometry. *Food Chemistry, 243*(March 2017), 420–427.

2. OBJECTIVES

2. Objectives

2.1 General Objective

The main objective of this thesis was to evaluate the potential use of near infrared spectroscopy as a tool for quality control in the cocoa powder industry, especially to quickly and non-destructively identify the presence and quantify the content of unwanted materials (cocoa shell) or adulterants (carob flour), as well as to predict the content of the main bioactive compounds of cocoa powders (methylxanthines and flavanols).

2.2. Specific Objectives

• To create a collection of representative cocoa samples with different origins and alkalization degrees, as well as different cocoa shell and carob flours, and to use these samples to create a battery of "adulterated" samples with known quantities of cocoa shell and carob flour.

• To measure the NIR spectra of all the samples and to correlate them with the added cocoa shell or carob flour contents, or with the methylxanthines and flavanols content.

• To develop and an HPLC method for the quantification of the methylxanthines and flavanols (such as catechin, epicatechin, caffeine and theobromine) present in cocoa powders characterized by different alkalization.

• To assess the influence of the alkalization level on the methylxanthines and flavanols content of the cocoa powders.

• To obtain classification models able to discriminate between samples of cocoa powders in categories related to the levels of cocoa shell (> 5 % and <5-40%) or carob flour content (cocoa powders, adulterated cocoa powders and carob flours), the article I evaluated by means of the sensitivity, specificity and the NER and the article II evaluated by their accuracy.

Objectives

• To obtain prediction models able to determine the level of presence of cocoa shell (0 to 40%), the percentage of adulteration with carob flour in cocoa powder and the flavanols and methylxanthines content of the cocoa powders, which will be evaluated by means of statistical parameters such as the coefficients of determination (R^2), the Root Mean Square Errors and the Ratio Prediction Deviation (RPD).

3. SCIENTIFIC CONTRIBUTIONS

3. Scientific Contributions

The thesis is divided into two different chapters:

a. ***Identification of cocoa components out of normative limits or non-declared ingredients in cocoa powders.***

The results included in this chapter have been published into 2 articles:

- *ARTICLE I. Fast detection of cocoa shell in cocoa powders by near infrared spectroscopy and multivariate analysis.*
- *ARTICLE II. Rapid fraud detection of cocoa powder with carob flour using near infrared spectroscopy*

b. ***Prediction of the content of bioactive compounds of cocoa powders.***

The results included in this chapter are now considered for their publication in 1 article:

- *Article III. Changes in methylxanthines and flavanols during cocoa powder processing and its quantification by near-infrared spectroscopy.*

3.1 CHAPTER 1. Identification of cocoa components out of normative limits or non-declared ingredients in cocoa powders.

3.1.1 *Fast detection of cocoa shell in cocoa powders by Near Infrared Spectroscopy and multivariate analysis*

Maribel Alexandra Quelal Vásconez,[a] María Jesús Lerma-García,[a] Édgar Pérez-Esteve,[a] Alberto Arnau-Bonachera,[b,c] José Manuel Barat,[a] Pau Talens,[a*]

[a]*Departamento de Tecnología de Alimentos, Universitat Politècnica de València. Camino de Vera s/n, 46022, Spain*
[b]*Institute for Animal Science and Technology, Universitat Politècnica de València - Universidad de Valencia. Departamento de Química Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain*
[c]*Biomedical Research Institute (PASAPTA-Pathology group), Veterinary School, Universidad Cardenal Herrera-CEU, CEU Universities, Av. Seminario s/n, 46113 Moncada, Valencia, Spain*

Abstract

Cocoa shell must be removed from the cocoa bean before or after the roasting process. In the case of a low efficient peeling process or the intentional addition of cocoa shell to cocoa products (i.e. cocoa powders) to increase the economic benefit, quality of the final product could be unpleasantly affected. In this scenario, the Codex Alimentarius on cocoa and chocolate has established that cocoa cake must not contain more than 5% of cocoa shell and germ (based on fat-free dry matter). Traditional analysis of cocoa shell is very laborious. Thus, the aim of this work is to develop a methodology based on near infrared (NIR) spectroscopy and multivariate analysis for the fast detection of cocoa shell in cocoa powders. For this aim, binary mixtures of cocoa powder and cocoa shell containing increasing proportions of cocoa shell (up to ca. 40% w/w based on fat-free dried matter) have been prepared. After acquiring NIR spectra (1100–2500 nm) of pure samples (cocoa powder and cocoa shell) and mixtures, qualitative and quantitative analysis were done. The qualitative analysis was performed by using principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA), finding that the model was able to correctly classify all samples containing less than 5% of cocoa shell. The quantitative analysis was performed by using a partial least squares (PLS) regression. The best PLS model was the one constructed using extended multiple signal correction plus orthogonal signal correction pre-treatment using the 6 main wavelengths selected according to the Variable Importance in Projection (VIP) scores. Determination coefficient of prediction and root mean square error of prediction values of 0.967 and 2.43, respectively, confirmed the goodness of the model. According to these results it is possible to conclude that NIR technology in combination with multivariate analysis is a good and fast tool to determine if a cocoa powder contains a cocoa shell content out of Codex Alimentarius specifications.

Article 1

1. Introduction

Cocoa powder is a cocoa bean (*Theobroma cacao. L*) derivative largely consumed around the world due to its capacity to give color, flavor and eating pleasure to a myriad of food preparations (Dico *et al*., 2018).

The obtaining of cocoa powder from cocoa beans follows different steps. First of all, beans must be peeled, starting with the peeling of the bean before or after a roasting process. During the same peeling, cocoa cotyledon must be separated from cocoa shell (12–20% of the cocoa seed), yielding fragments of cotyledon, called nibs (Okiyama, Navarro, & Rodrigues, 2017). During the shelling step, shell should be perfectly separated, removing large parts of shells and leaving nib particles practically unbroken (Beckett, 2009). The performance of this proce- dure is very relevant since the presence of cocoa shell in cocoa beans derivatives (cocoa liquor, cocoa powder or chocolate) adversely affects the final product quality (Mendes & Lima, 2007). Concretely, it can have an influence in some characteristics of the final product such as the flavor or taste; it can also be responsible of off-flavors. Additionally, fiber content in cocoa shell is really high. Thus, it can be a problem for the grinding process, causing equipment abrasion in some cases.

Bearing this in mind it is not surprising that shell content in cocoa powders is a quality parameter to be controlled. Concretely, the Codex Alimentarius establishes a maximum amount of 5% of cocoa shells in cocoa cake (based on fat-free dry matter) (Codex Alimentarius, 2016).

Analysis of cocoa shell in cocoa products might be done following the AOAC 968.10 or the 970.23 methods (Codex Alimentarius, 2016). The first method, called spiral vessel count consists of counting the spiral vessels in a defatted, grinded and digested sample with the help of a microscope adjusted to mold counting (field of view 1.382mm at 100x) (AOAC, 1084). The second method, called stone cell count, consists of microscope assisted counting the stone cells present in the samples after a really laborious preparation (AOAC, 1984).

Since those methods are really arduous, recent attempts to develop alternative methods have been done. Researchers from the Nestlé Research Center proposed a gas-liquid chromatography procedure based on the detection of fatty acid tryptamides (FATs) in the sample, since FATs are compounds more abundant in cocoa shells than in other parts of cocoa seed. This work, carried out with only cocoa originating from the Ivory Coast, demonstrated that it might be an appropriate tool for the determination and prediction of the shell content in cocoa liquor (Hug, Golay, Giuffrida, Dionisi, & Destaillats, 2006). In another work, Yang *et al.* (2015) proposed the employment of polysaccharide fingerprint established by high performance liquid chromatography followed by principal component analysis to identified cocoa powders adulterated with cocoa or other plant shells such as chestnut, longan, peanut, etc. However, only cocoa powders containing cocoa or other plant shell percentages higher than 15 and 10%, respectively, were detected using this methodology. Therefore, even when these methodologies (determination of FATs, HPLC polysaccharide fingerprint, etc.) are more sensible, accurate and faster than the methods proposed by the Codex Alimentarius, their use as routine techniques for shell content determination still have certain limitations such as the limit of detection or the fact that they need sample preparation, require specialized personnel and they are destructive. To avoid these drawbacks common in traditional chemical analysis techniques, recent attempts on developing accurate and sensi- tive analytic techniques based on near infrared (NIR) spectroscopy have been done. Due to the ability of NIR spectroscopy to provide a spectrum that acts as a 'fingerprint' distinctive of a particular sample, this technology is now widely used as a successful quality control tool (Lerma-García, Cortés, Talens, & Barat, 2018). Concretely, in the cocoa sector NIR spectroscopy has been employed for the prediction of majority (moisture, carbohydrate, fat, protein) or minority functional compounds (theobromine, catechin, organic acids, etc.) (Krähmer *et al.*, 2015; Veselá *et al.*, 2007; Álvarez *et al.*, 2012) as well as for quality control (dis- crimination of cocoa beans according to geographical origin, prediction of cocoa powder adulterations, etc) (Quelal *et al.*, 2018; Teye, Huang, Dai, & Chen, 2013).

Article 1

In this scenario, the goal of this work is the fast determination of cocoa shell content in cocoa powders in concentrations higher than the limit established by the Codex Alimentarius (5%) by means of NIR spectroscopy and a multivariate analysis.

2. Materials and methods

2.1 Cocoa powder and shell Samples

A total of 20 natural cocoa powders and 2 cocoa shells, gently provided by Olam Food Ingredients (Cheste, Spain) or purchased in the market from different origins (Ghana, Ivory Coast, Cameroon, Peru and Indonesia) were employed in this study. In order to predict the presence of cocoa shell in cocoa powders using partial least squares (PLS), binary mixtures containing cocoa powder and cocoa shell were prepared. The mixtures contained percentages of cocoa shells in cocoa powder (based on fat-free dry matter) from ca. 2.5–40%. Percentages higher than 40% were not considered since over this percentage the presence of cocoa shell is sensory evident. To improve the robustness of the PLS model, all 20 cocoa powder samples (coming from different origins and obtained after different processings) were randomly selected to perform a total of 12 binary mixtures for each percentage (2.5, 5, 7.5, 10, 20 and 40%), in which both cocoa shell samples were also considered. Thus, a total of 72 mixtures were obtained. Once all mixtures were prepared, they were poured in hermetic plastic containers and stored at 20 ± 2 °C under dark conditions until use.

2.2 NIR spectra acquisition

The 94 samples (20 cocoa powders, 2 cocoa shells and 72 binary mixtures) were measured with a spectrophotometer FOSS NIR 5000 (Silver Spring, MD, USA). A uniform thickness and surface were secured during spectra scanning using a device with 380 mm of diameter and 1cm of thick with a quartz windows which was filled with 5 g of sample. The spectrophotometer gives the measurements in relative

94

absorbance units (log 1/R), which could be correlated with chemical constituents (Liu, Sun, & Ouyang, 2010; Martens, Nielsen, & Engelsen, 2003). Each sample was scanned 32 times in a range comprised between 1100 and 2500 nm at 2 nm intervals (700 points). The samples were measured twice and no differences between them were found.

2.3 Statistical analysis

Spectral data were pre-treated and analysed using qualitative and quantitative models by means of the chemometric software Unscrambler v10.5 (CAMO Software AS, Oslo, Norway).

The PCA model was performed using raw data to identify different sample groups and to find and remove defective outliers (Adnan, Hörsten, Pawelzik, & Mörlein, 2017; Bro & Smilde, 2014).

The PLS was performed in order to predict the presence of cocoa shell in the cocoa powders and the PLS-DA (Berrueta, Alonso, & Héberger, 2007; Prats-Montalbán, Jerez-Rozo, Romañach, & Ferrer, 2012), was constructed to evaluate its capability in classifying samples according to the following categories: cocoa powders containing less than 5% cocoa shell (w/w), and cocoa powders containing from 5 to 40% cocoa shell (w/w).

Both analyses were performed using the pre-treated spectra. The spectral pre-treatments tried included extended multiple signal correction (EMSC) (Martens *et al.*, 2003), standard normal variation (SNV), 2nd derivative with the Savitzky-Golay (S-G), orthogonal signal correction (OSC) and combinations of all of them with OSC.

To construct both PLS and PLS-DA models, two data matrices were used. The first one employed for the PCA and PLS model construction, contained the spectra of all samples (N=94) and the same 700 X- variables. In this case, all individual cocoa shell percentages were considered as Y-variable. The second matrix, employed for the PLS-DA model construction, included the spectra of 92 samples

(in which the spectra of cocoa shells were not considered since the considered categories were cocoa shell contents below 5% and between 5 and 40%) and 700 predictors or X-variables (wavelengths), and also a dependent Y-variable containing the 2 categories previously described (<5% and 5–40% cocoa shell based on fat-free dry matter, w/w).

For both, PLS-DA and PLS models construction, the use of all spectra wavelengths was considered, jointly with the use of the most important wavelengths. The PLS and the score of Variable Importance in Projection (VIP) were combined together for these selection (Botelho, Reis, Oliveira, & Sena, 2015).

To select the optimal factor number and to avoid the over-fitting of both PLS and PLS-DA models, leave-one-out cross-validation was used using 70% of the data, which were randomly selected. The remaining 30% of the data were used as an external validation set.

PLS models accuracy was evaluated by the required number of latent variables (LVs), the coefficient of determination of calibration (R^2_C), RMSEC, the coefficient of determination of cross-validation (R^2_{CV}), RMSECV, the coefficient of determination for prediction (R^2_P), the root mean square error of prediction (RMSEP), the ratio of prediction deviation (RPD, which is calculated as ratio between the standard deviation of reference values in training set and RMSEP) and the bias value (which establishes the difference between experimental values and NIR predictions). Bias value can be positive (overestimating) or negative (underestimating), indicating values near to zero a minimum deviation from experimental and predicted values (Cantor, Hoag, Ellison, Khan, & Lyon, 2011).

On the other hand, the number of latent variables (LVs) for the PLS-DA model was determined by the low value of the root mean square error of calibration (RMSEC), and the root mean square error of leave-one-out cross-validation (RMSECV) (Botelho *et al.*, 2015). The PLS-DA classification performance was evaluated by sensitivity, specificity and by the non-error rate (NER). Sensitivity is the model ability related to a correct classification of the samples with different levels of cocoa shell content. The model capacity to correctly determine the samples

96

which not correspond to the class and correctly refuse them is the specificity (Almeida, Fidelis, Barata, & Poppi, 2013). The non-error rate (NER) is the average of the sensitivities of the different categories (Manfredi, Robotti, Quasso, Mazzucco, Calabrese, & Marengo, 2018).

3. Results and discussion

3.1 Cocoa powder and shell spectra, pre-treated spectra and PCA analysis

The mean raw spectra of cocoa powders, cocoa shells and binary mixtures of them at different percentages are shown in **Figure. 1a**. As shown in this figure, the main bands observed appeared at 1470, 1930 and 2130 nm, although other bands at 1730, 2310 and 2350 nm were also evidenced. Although all spectra have a similar pattern of absorbance, the relative absorbance of these bands is different for the different types of samples: cocoa shell is characterized by the highest relative absorbance, which decreased when the content of cocoa shell in the samples decreased. The signal at 1470 nm correspond to the firs overtone of O-H and N-H stretching which is associated with a $CONH_2$ structure (peptide) and related to a protein (Osborne, Fearn, & Hindle, 1993).

The signal at 1930 nm is related with asymmetric stretching and rocking of water, weakly bounded water, proteins, and aromatics (Veselá *et al*., 2007), while the wavelength at 2130 nm can be assigned to N–H combination bands ($CONH_2$) (Ribeiro, Ferreira, & Salva, 2011). On the other hand, the band at 1730 nm could be assigned to the first overtone of C-H (Ribeiro *et al*., 2011), while 2310 and 2350 nm are mostly related to stretching and rocking vibrations of CH2 of polysaccharides (Veselá *et al*., 2007).
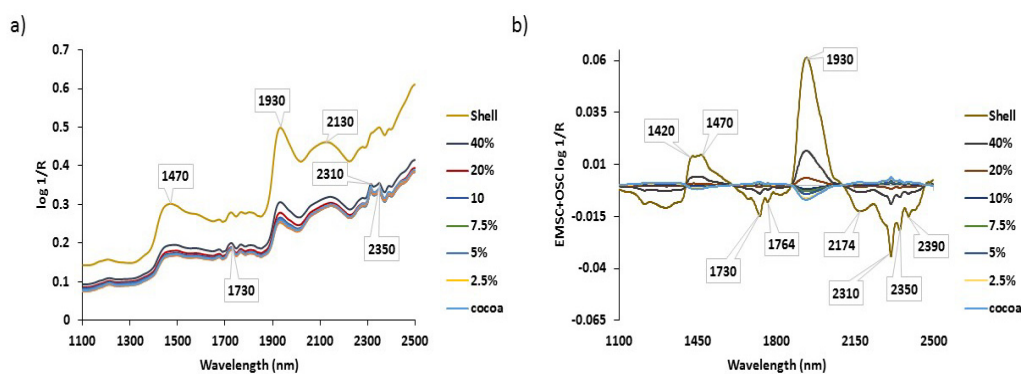
Figure 1. Mean spectra of cocoa powders and shells and mixtures of them at different percentages from (a) raw and (b) pre-treated with EMSC-OSC spectra.

The mean spectra obtained after the application of the EMSC-OSC pre-treatment is shown in Figure. 1b. In this case, the principal wavelengths were 1420, 1470, 1730, 1764, 1930, 2174, 2310, 2350 and 2390 nm. Most of the bands have been previously described, while the other ones could be attributed to the first overtones of symmetric and anti-symmetric C-H stretch vibration (CH_2-groups) (1764 nm) (Krähmer *et al.*, 2015), to a combination of C-H (2174 nm) (Ma *et al.*, 2017) and to the combination of C-H stretch and C-H deformation modes (2390 nm) (Wang *et al.*, 2018).

In order to have a more precise idea about the relation between samples and variables a PCA model, a non-supervised method was performed with the raw spectra data to identify possible sample groupings. The score plot of the two first principal components (PCs) is shown in **Figure. 2**. A total of 98% of the variance is explained by these two first PCs (87 and 11% for PC_1 and PC_2, respectively). Along PC_1, cocoa shell samples are clearly separated from the remaining ones, in which any clear tendency was observed, although samples containing high cocoa shell percentages (40% w/w) seemed to be located closer to the PC_1 values of cocoa shell. According to the X-loading values (data not shown), the wavelengths with higher weights were 1930, 1420 and 1470 nm for the PC_1 and for the PC_2 were 1644, 1326,

2146, 2310 and 2350 nm. Some of these peaks (1930, 1470, 2310 and 2350 nm) matched with the main peaks observed in raw spectra, which have been previously mentioned. The other bands corresponded to the fist overtone of the hydroxyl and amino groups (1420 nm) and first overtone of C-H (1644 nm) (Ribeiro *et al*., 2011), the second overtone of C-H (1326 nm) (Ma, Wang, Chen, Cheng, & Lai, 2017) and the combination of C-C and C-H stretching (2146 nm) (Workman, & Weyer, 2008).
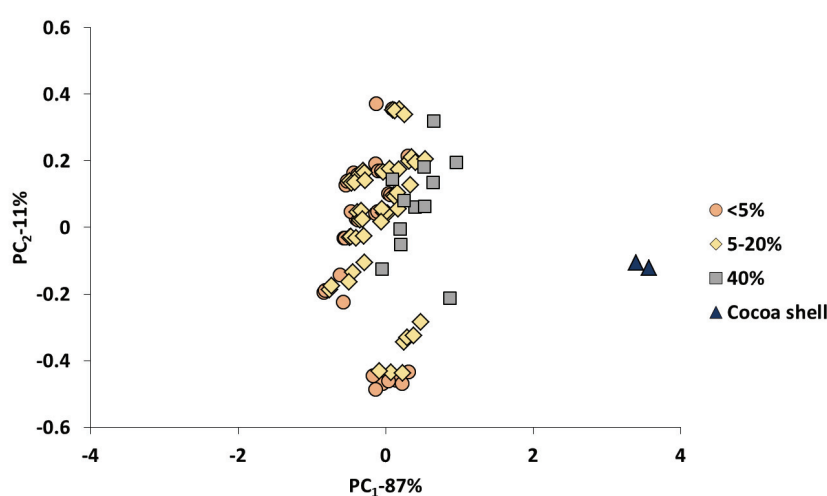

Figure 2. PCA score plot of the two first PCs showing the distribution of all the samples considered in this study. Samples were labelled as follows: cocoa shell content < 5%, comprised between 5 and 20%, 40% and pure cocoa shells.

3.2 *Prediction of the added cocoa shell percentage in cocoa powders by PLS*

A total of 8 PLS models using all the available wavelengths (700) as variables, one for each pre-treatment considered in the study, were performed. The results obtained are summarized in Table 1. At the sight of the results, the best PLS model was the one constructed using the EMSC+OSC pre-treatment. In order to reduce the high dimensionality of the spectral data, the most important wavelengths were selected according to the VIP scores (**Figure. 3**). These VIP scores determine the significance of each variable in the projection used by a given PLS model by means of their coefficients in every component, jointly with the significance of each

component in regression (Botelho *et al*., 2015). As it could be observed in Figure. 3, the most important variables are wavelengths at 1930, 1420 and 1470 nm at positive values of LV_1, and 2310, 2350 and 1730 nm at negative values of LV_1. These wavelengths are mostly the same previously mentioned in both raw and pre-treated spectra, which demonstrated their importance in cocoa shell content prediction. Most of these wavelengths have been previously described in literature in the prediction of several compounds (such as fat, carbohydrates, polysaccharides, moisture, polyphenols, etc.) of cocoa beans and derived products (Huang *et al*., 2014; Krähmer *et al*., 2015; Quelal-Vásconez, Pérez-Esteve, Arnau-Bonachera, Barat, & Talens, 2018; Veselá *et al*., 2007).
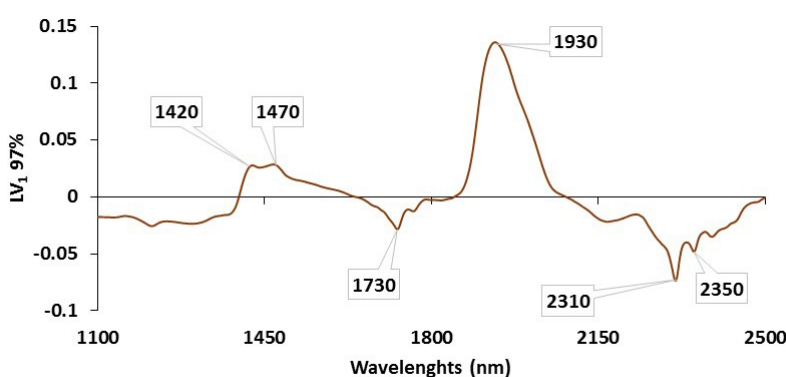
Figure 3. Variable importance in projection (VIP) scores of the PLS model constructed to predict cocoa shell percentages.

Using the EMSC+OSC pre-treatment and the six wavelengths obtained in the VIP scores as variables, another PLS model was constructed. The results obtained for this model are also shown in Table 1. Compared to the best model obtained with the same pre-treatment but using all the available wavelengths, this model is less complex although all the other parameter values are very similar.

Table 1: Results of the PLS models constructed for predicting cocoa shell percentage using different pre-treatments and different number of wavelengths with a calibration and validation sets.

Pretreatment	#W	#LV	Calibration		Cross-validation		Validation			
			R^2_C	RMSEC	R^2_{CV}	RMSECV	R^2_P	RMSEP	Bias	RPD
Raw data	700	7	0.908	3.68	0.694	6.83	0.930	3.52	0.351	3.46
EMSC	700	7	0.936	3.06	0.857	4.64	0.941	3.24	0.095	3.77
SNV	700	7	0.931	3.18	0.862	4.55	0.940	3.27	0.057	3.72
2nd Der. (S-G)	700	7	0.967	2.20	0.936	3.09	0.955	2.96	-0.021	4.11
OSC	700	1	0.990	1.20	0.989	1.25	0.851	5.16	-0.059	2.36
EMSC-OSC	*700*	1	0.974	1.92	0.973	2.01	0.967	2.41	0.204	5.06
SNV+OSC	700	1	0.978	1.79	0.976	1.89	0.967	2.55	-0.278	4.77
2nd Der. (S-G)+OSC	700	3	0.944	2.85	0.942	2.96	0.939	3.33	-0.104	3.66
EMSC-OSC	*6*	*1*	*0.975*	*1.91*	*0.973*	*2.01*	*0.967*	*2.43*	*0.195*	*5.03*

#W = number of wavelengths used to construct de model; #LV = latent variables; R2 = determination coefficient; RMSEC = Root mean square error of calibration; RMSECV = Root mean square error of cross-validation; RMSEP = Root mean square error of prediction; RPD = Ratio prediction deviation; EMSC = Extended multiple scatter correction; 2nd Der. (S-G) = Second derivative and Savitzky Golay smoothing, SNV = Standard Normal Variate, OSC = Orthogonal signal correction.

The plot representing the predicted versus the measured cocoa shell percentages of the prediction set samples constructed with PLS data of the model constructed using the 6 wavelengths as variables is shown in Figure. 4. A good linear fit due to the closer relationship between the reference values and the NIR spectra is observed, displaying the reliability and accuracy of the NIR in determining the percentage of cocoa shell present in the cocoa powders.
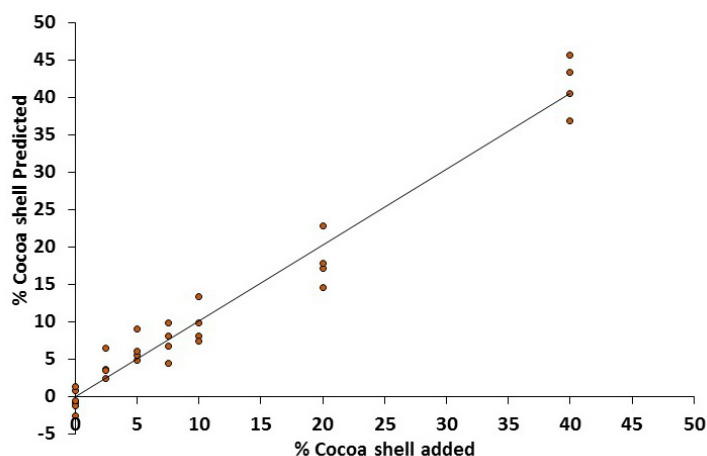
Figure 4. Predicted versus measured cocoa shell percentages by PLS model constructed using the 6 main wavelengths in the prediction set.

3.3 Classification of cocoa powder samples according to the added level of cocoa shell

Since PCA is a non-supervised method, and it is not possible to observe a clear separation between the different sample categories, a supervised discriminant model, PLS-DA, was next constructed using all the available wavelengths (700) and the EMSC-OSC pre-treatment. The best model was obtained with 2 LVs with RMSEC and RMSECV values of 0.24 and 0.28, respectively, with most of the variability explained by the LV_1 (72%).

Next, using the 6 most relevant wavelengths as variables, another PLS-DA model was constructed. The discriminant plot obtained using the two LVs for the classification of samples according to the different categories is shown in Figure. 5. As it can be observed in this figure, separation between the two categories is achieved along LV_1, with negative scores related to the samples containing < 5% cocoa shell, and positive scores related to samples containing 5-40% cocoa shell.
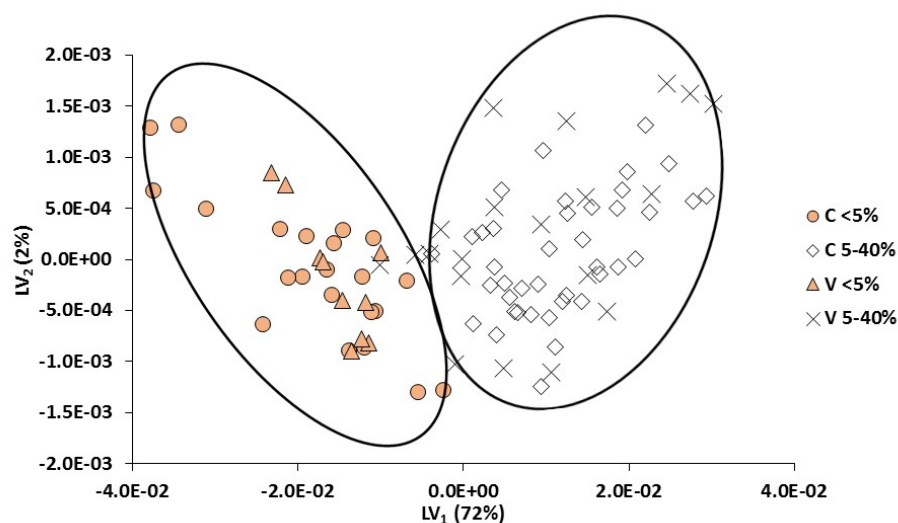
Figure 5. PLS-DA discriminant plot constructed using the two first LVs to classify cocoa powders according to the following categories: cocoa shell content < 5% and cocoa shell content comprised between 5 and 40%. Both calibration (C<5% and C 5-40%) and external validation (V<5% and V 5-40%) set samples have been included and represented with different symbols.

Once constructed, the model was validated with the external validation set samples. The results obtained for both calibration and external validation sets for this model are included in Table 2. As it can be observed in the confusion table for the calibration samples, all samples were correctly classified. On the other hand, for the external validation set, all samples of the <5% category were correctly classified, while 3 samples of the 5-40% category were misclassified. Even if the number of misclassified samples is very low, it should be highlighted that all the "misclassified samples" corresponded to samples containing a 5% cocoa shell (based on fat-free dry matter), which is the limit established by the Codex Alimentarius, and thus the borderline of both categories. Next, the PLS-DA classification performance was evaluated by the sensitivity, specificity and NER values, which are also included in Table 2. Taking into account the values reported and the comments previously mentioned, it could be concluded that the PLS-DA model constructed is able to

reliable discriminate between samples containing cocoa shell percentages below and upper 5%.

Table 2. Confusion table, sensitivity (SENS), specificity (SPEC) and non-error prediction rates (NER) of the PLS-DA model constructed with variable selection to discriminate cocoa powders into two categories: cocoa powders with < 5% and between 5-40% cocoa shell.

	Category		# Samples	SENS (%)	SPEC (%)	NER (%)
	<5%	5-40%				
Calibration set samples						
<5%	22	0	22	100	100	100
5-40%	0	40	40	100	100	
	22	40	62			
External validation set samples						
<5%	10	0	10	100	85	92.5
5-40%	3	17	20	85	100	
	13	17	30			

4. Conclusions

NIR spectroscopy in combination with PLS and PLS-DA statistical models has been shown to be a rapid and effective method to determine cocoa shell content in cocoa powders. Using a PLS analysis, it was possible to quantify the percentage of cocoa shell present in cocoa powders. The best PLS prediction model was constructed using the 6 main wavelengths (1420, 1470, 1730, 1930, 2310 and 2350) selected according to the VIP scores, obtaining 1 LV with R^2_C and R^2_{CV} of 0.975 and 0.973, respectively, and RMSEC and RMSECV of 1.91 and 2.01, respectively. Regarding the validation samples, R^2_P was 0.967 while RMSEP was 2.43, confirming the goodness of the model. On the other hand, the PLS-DA analysis show that 92.5% of the validation set samples were correctly classified into two groups: samples with a shell content lower than 5% (considered the acceptance limit in cocoa

powders by the Codex Alimentarius) and shell contents between 5 and 40%. These results indicate that this technology is therefore an important tool for cocoa producers and clients, who will be able to discriminate among samples in or out specifications, avoiding the use of destructive techniques that require a complex preparation of the sample or techniques that imply an important expense for the company.

Acknowledgements

References

Adnan, A., Hörsten, D. von, Pawelzik, E., & Mörlein, D. (2017). Rapid Prediction of Moisture Content in Intact Green Coffee Beans Using Near Infrared Spectroscopy. *Foods, 6*(6), 38.

Almeida, M. R., Fidelis, C. H. V, Barata, L. E. S., & Poppi, R. J. (2013). Classification of Amazonian rosewood essential oil by Raman spectroscopy and PLS-DA with reliability estimation. *Talanta, 117*, 305–311.

Álvarez, C., Pérez, E., Cros, E., Lares, M., Assemat, S., Boulanger, R., & Davrieux, F. (2012). The use of near infrared spectroscopy to determine the fat, caffeine, theobromine and (−)-epicatechin contents in unfermented and sun-dried beans of Criollo cocoa. *Journal of Near Infrared Spectroscopy, 20*(2), 307–315.

AOAC, 1984. AOAC, Official methods of analysis. Association of Official Analytical Chemists. 14th ed., Arlington, VA.

Article 1

AOAC. (2006). AOAC Official Method 968.10 Spiral Vessel Count of Cacao Products, 2-3.

Berrueta, L. A., Alonso, R. M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. Journal of Chromatography A, 1158(1–2), 196–214.

Beckett, S. T. (2009). Industrial chocolate manufacture and use. (4th ed.). Hoboken, NJ: Wiley-Blackwell.

Botelho, B. G., Reis, N., Oliveira, L. S., & Sena, M. M. (2015). Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA. *Food Chemistry, 181*, 31–37.

Bro, R., & Smilde, A. K. (2014). Principal component analysis. Analytical *Methods, 6*(9), 2812–2831.

Cantor, S. L., Hoag, S. W., Ellison, C. D., Khan, M. A., & Lyon, R. C. (2011). NIR Spectroscopy Applications in the Development of a Compacted Multiparticulate System for Modified Release. *AAPS PharmSciTech, 12*(1), 262–278.

Codex Alimentarius (2016). Standard for cocoa (cacao) mass (cocoa/chocolate liquour) and cocoa cake. Codex Stan 141-1983.

Dico, G. M. L., Galvano, F., Dugo, G., D'ascenzi, C., Macaluso, A., Vella, A., & Ferrantelli, V. (2018). Toxic metal levels in cocoa powder and chocolate by ICP-MS method after microwave-assisted digestion. *Food Chemistry, 245*, 1163–1168.

Huang, X. Y., Teye, E., Sam-Amoah, L. K., Han, F. K., Yao, L. Y., & Tchabo, W. (2014). Rapid measurement of total polyphenols content in cocoa beans by data fusion of NIR spectroscopy and electronic tongue. *Analytical Methods, 6*(14), 5008–5015.

Hug, B., Golay, P. A., Giuffrida, F., Dionisi, F., & Destaillats, F. (2006). Development of a gas−liquid chromatographic method for the analysis of fatty acid tryptamides in cocoa Products. *Journal of Agricultural and Food Chemistry, 54*(9), 3199–3203.

106

Krähmer, A., Engel, A., Kadow, D., Ali, N., Umaharan, P., Kroh, L. W., & Schulz, H. (2015). Fast and neat – Determination of biochemical quality parameters in cocoa using near infrared spectroscopy. *Food Chemistry, 181*, 152–159.

Lerma-García, M. J., Cortés, V., Talens, P., & Barat, J. M. (2018). Variety discrimination of fruits, edible plants, and other foodstuffs and beverages by Infrared Spectroscopy. In J. Lopes, & C. Sousa (eds.), *Comprehensive Analytical Chemistry volume 80* (pp. 127–163). Oxford: Elsevier.

Liu, Y., Sun, X., & Ouyang, A. (2010). Nondestructive measurement of soluble solid content of navel orange fruit by visible-NIR spectrometric technique with PLSR and PCA-BPNN. LWT - *Food Science and Technology, 43*(4), 602–607.

Ma, H. L., Wang, J. W., Chen, Y. J., & Lai, Z. T. (2017). Rapid authentication of starch adulterations in ultrafine granular powder of Shanyao by near-infrared spectroscopy coupled with chemometric methods. *Food chemistry, 215*, 108-115.

Manfredi, M., Robotti, E., Quasso, F., Mazzucco, E., Calabrese, G., & Marengo, E. (2018). Fast classification of hazelnut cultivars through portable infrared spectroscopy and chemometrics. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy, 189*, 427–435.

Martens, H., Nielsen, J. P., & Engelsen, S. B. (2003). Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures. *Analytical Chemistry, 75*(3), 394–404.

Mendes, F. A. T., & Lima, E. L. (2007). Perfil Agroindustrial do Processamento de Amêndoas de Cacau em Pequena Escala no Estado do Pará. Belém. *SEMBRAE/PA*.

Osborne, B. G., Fearn, T., & Hindle, P. H. (1993). Practical NIR spectroscopy with applications in food and beverage analysis. Longman Scientific & Technical, Wiley, Harlow, Essex, England, New York.

Okiyama, D. C., Navarro, S. L., & Rodrigues, C. E. (2017). Cocoa shell and its compounds: Applications in the food industry. *Trends in Food Science & Technology, 63*, 103–112.

Article 1

Prats-Montalbán, J. M., Jerez-Rozo, J. I., Romañach, R. J., & Ferrer, A. (2012). MIA and NIR Chemical Imaging for pharmaceutical product characterization. *Chemometrics and Intelligent Laboratory Systems, 117*, 240–249.

Teye, E., Huang, X., Dai, H., & Chen, Q. (2013). Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 114*, 183–189.

Quelal-Vásconez, M. A., Pérez-Esteve, É., Arnau-Bonachera, A., Barat, J. M., & Talens, P. (2018). Rapid fraud detection of cocoa powder with carob flour using near infrared spectroscopy. *Food Control, 92*, 183–189.

Ribeiro, J. S., Ferreira, M. M. C., & Salva, T. J. G. (2011). Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy. *Talanta, 83*(5), 1352–1358.

Veselá, A., Barros, A. S., Synytsya, A., Delgadillo, I., Čopíková, J., & Coimbra, M. A. (2007). Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder. *Analytica Chimica Acta, 601*(1), 77–86.

Wang, J., Zhang, X., Sun, S., Sun, X., Li, Q., & Zhang, Z. (2018). Online determination of quality parameters of dried soybean protein–lipid films (Fuzhu) by NIR spectroscopy combined with chemometrics. *Journal of Food Measurement and Characterization*, https://doi.org/10.1007/s11694-018-9762-z.

Workman Jr, J., & Weyer, L. (2007). Practical guide to interpretive near-infrared spectroscopy. Taylor and Francis Group, Boca Raton, FL: CRC Press.

Yang, W., Hu, M., Chen, S., Wang, Q., Zhu, S., Dai, J., & Li, X. (2015). Identification of Adulterated Cocoa Powder Using Chromatographic Fingerprints of Polysaccharides Coupled with Principal Component Analysis. *Food Analytical Methods, 8*(9), 2360–2367.

3.1.2 Rapid fraud detection of cocoa powder with carob flour using near infrared spectroscopy

Maribel Alexandra Quelal Vásconez,[a] Édgar Pérez-Esteve,[a] Alberto Arnau-Bonachera,[b,c] José Manuel Barat,[a] Pau Talens,[a*]

[a]Departamento de Tecnología de Alimentos, Universitat Politècnica de València. Camino de Vera s/n, 46022, Spain

[b]Institute for Animal Science and Technology, Universitat Politècnica de València - Universidad de Valencia. Departamento de Química Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain

[c]Biomedical Research Institute (PASAPTA-Pathology group), Veterinary School, Universidad Cardenal Herrera-CEU, CEU Universities, Av. Seminario s/n, 46113 Moncada, Valencia, Spain

Abstract

Cocoa powder is a highly valuable global product that can be adulterated with low-cost raw materials like carob flour as small amounts of this flour would not change the color, aroma and taste characteristics of the final product. Rapid methods, like NIR technology combined with multivariate analysis, are interesting for such detection. In this work, unaltered cocoa powders with different alkalization levels, carob flours with three different roasting degrees, and adulterated samples, prepared by blending cocoa powders with carob flour at several proportions, were analysed. The diffuse reflectance spectra of the samples of 1100 - 2500 nm were acquired in a Foss NIR spectrophotometer. A qualitative and a quantitative analysis were done. For the qualitative analysis, a principal component analysis (PCA) and a partial least squares discriminant analysis (PLS-DA) were performed. Good results (100% classification accuracy) were obtained, which indicates the possibility of distinguishing pure cocoa powders from adulterated samples. For the quantitative analysis, a partial least squares (PLS) regression analysis was performed. The most robust PLS prediction model was obtained with one factor (LV), a coefficient of determination for prediction (RP2) of 0.974 and a root mean square error of prediction (RMSEP) of 3.2% for the external set. These data allowed us to conclude that NIR technology combined with multivariate analysis enables the identification and determination of the amount of natural cocoa powder present in a mixture adulterated with carob flour.

Keywords: Cocoa powder, adulteration, carob flour, NIR, PCA, PLS.

Article 2

1. Introduction

Cocoa powder, thanks to its characteristic and pleasant flavour and aroma, is one of the most valued commodities around the world (Bonvehí, 2005). Among its applications in the food industry, the formulation of beverages, confectionery, bakery and pastry products stands out (Shankar, Levitan, Prescott, & Spence, 2009). Apart from flavour and aroma, cocoa is highly appreciated as a natural colouring agent, partly because of the current tendency to restrict the use of artificial colors.

During cocoa processing, it is possible to modify cocoa color and aroma through roasting and/or alkalization processes. Roasting consists of exposing cocoa beans to temperatures of 130–150°C for 15–45 min. It is used to inactivate microorganisms and to develop the characteristic brown color, mild aroma and texture of commercial natural beans (Bonvehí, 2005, Krysiak, 2006; Afoakwa, Budu, Mensah-Brown, Felix & Ofosu-Snsah, 2014). Alkalization is an optional operation to reduce acidity, bitterness and astringency, and to darken cocoa's color. This procedure involves using an alkali (generally potassium carbonate) in combination with oxygen, water and high temperatures. These extreme conditions provoke, among others, Maillard reactions and polyphenol oxidations and polymerizations, which end up with flavor and color modifications from light brown (natural) to red, dark brown or extremely black (Miller *et al*., 2008; Li *et al*., 2012).

In recent years, the demand for cocoa powder has increased and its supplies have tightened, thus its price has steadily grown (Fadel, Mageed, Samad, & Lotfy, 2006). Consequently, there is a demand to develop cocoa substitutes. Some studies suggest that cocoa-like aromas can be found in roasted carobs (Arrighi, Hartman & Ho, 1997). Carob pods are characterized by a high sugar content (around 50%), composed essentially of sucrose. This high sugar content favors the same chemical reactions that occur during the roasting and alkalization of cocoa: caramelization of high sugar content and Maillard reactions between amino acids and sugars (Fadel *et al*., 2006). In this way, toasted carob can provide similar aromas to cocoa.

Bearing in mind this striking aromatic and visual similarity between carob flour (natural or toasted) and cocoa (natural or alkalized), some traders have seen that selling carob (average price of 940 US$/tonne) as cocoa (1945 US$/tonne), by omitting this substitution, is a profitable option to increase their benefits (ICCO, 2017). However, this deliberate, intentional and undeclared substitution of one product for another with a lower price is food fraud that not only affects producers and consumers, but also the physico-chemical properties of the manufactured product. Some studied examples comprise milk chocolates and chocolate cakes, in which some percentages of cocoa powder have been substituted for carob flour (Salem & Ohaad Fahad, 2012; Rosa, Tessele, Prestes, Silveira, & Franco, 2015).

To detect food adulteration, the three most widespread technologies are liquid chromatography, infrared spectroscopy and gas chromatography (Moore, Spink, & Lipp, 2012). Liquid and gas chromatography analyses need long sample preparation times, method optimization, and high-cost materials and reagents, while infrared spectroscopy is fast, reliable, less expensive and a chemical-free alternative (Ellis *et al.*, 2012). Near infrared spectroscopy (NIR) is an infrared spectroscopy type characterized by recording reflectance or transmittance spectra within the region from 750 nm to 2500 nm. These spectra act as a 'fingerprint' that is characteristic of a particular sample molecule and allows its identification. Some examples of using NIR and multivariate analyses in the cocoa sector include the prediction of basic food components, such as moisture, carbohydrate, fat, protein, theobromine and catechin and total polyphenol content (Veselá *et al.*, 2007; Álvarez *et al.*, 2012; Huang *et al.*, 2014). In other sectors, NIR in combination with a multivariate analysis has been employed to detect starch in onion powders, acid whey, starch, maltodextrin in skim powder milk, Sudan dyes in chili powders, and talcum powder in teas (Lohumi, Lee, Lee, & Cho, 2014; Capuano, Boerrigter-Eenling, Koot, & van Ruth, 2015; Haughey, Galvin-King, Ho, Bell, & Elliott, 2015; Li, Zhang, & He, 2016).

Article 2

In this context, the aim of this work was the rapid detection of the adulteration of cocoa powders, regardless of their alkalization level, with carob flours by applying NIR and a multivariate analysis.

2. Materials and methods

2.1 Raw materials

In order to analyse a good representative set of samples of the variability in commercial cocoa and carob flour, cocoa powders with different alkalization levels (n=12), as well as carob flour powders with three different roasting degrees (n=6) were used in this study. The samples used were natural cocoa (NC), lightly alkalized cocoa (LAC), medium alkalized cocoa (MAC), strong alkalized cocoa (SAC), light carob flour (LCF), medium carob flour (MCF) and dark carob flour (DCF).

OLAM Food Ingredients, Spain (Cheste, Valencia), kindly donated cocoa powders. Carob flour powders were bought from a local specialized supermarket. Raw samples were placed inside a glass container and stored in a dry dark atmosphere until were used.

2.2 The physico-chemical characterization of raw materials

Each of the raw samples was characterized according to their extractable pH value and extrinsic color. All measurements were taken in triplicate. For extractable pH determination, the process described in The Zaan Manual (Olam, 2017) was followed. For that purpose, 10g of cocoa powder were suspended in 90 mL of boiling distilled water and stirred. After decreasing temperature to 20-25ºC in a cold bath, sample pH was measured with a digital pH-meter (Crison Instruments, S.A., Barcelona, Spain) previously calibrated with 3 buffer solutions: pH 4.01, pH 7.0 and pH 9.21 (T=25 ºC). According to pH value, samples were classified in four different

categories: natural cocoa powders (5< pH<6), light alkalized (6<pH<7.2), medium alkalized (7.2< pH<7.6) and strong alkalized powders (pH >7.6) (Miller *et al.*, 2008).

To determine the extrinsic color, a cocoa powder sample was placed in a methacrylate cuvette by unifying the degree of compaction through small successive shocks. Color was measured in a spectrocolorimeter Minolta CM 3600D (Tokyo, Japan). Reflectance spectra (between 400 - 700 nm) were used to obtain color coordinates L*, a* and b* for D65 illuminant and 10° observer. Hue (h^*) and chroma (C^*) were estimated by Equation 1 and Equation2, respectively.

$$h^* = arctg \frac{b^*}{a^*} \qquad (1)$$

$$C^* = \sqrt{a^{*^2} + b^{*^2}} \quad (2)$$

2.3 Preparing adulterated samples

In this study, two batches of 234 samples composed of 12 unaltered cocoa powders, 6 carob flours, and 216 adulterated samples, were used. The adulterated samples were prepared by blending the 12 cocoa powders with the 6 different carob flours at different proportions. For all the 72 possible cocoa-carob combinations, three different levels of adulteration were prepared: low adulteration LA (0-20%), medium adulteration MA (20-40%) and high adulteration HA (40-60%). The upper limit (60%) was set by considering that above this concentration, adulteration would become evident due to the characteristic carob aroma (Cantalejo, 1997). The specific adulteration percentage at a given level was determined randomly from a uniform distribution (each adulteration percentage had the same probability of being selected), following the Latin Hypercube Strategy (LHS) (Helton & Davis, 2003). The adulterated samples were placed in a glass container and stored in a dry dark atmosphere until used.

Article 2

2.4 Collecting near-infrared spectra

All samples were scanned in a FOSS NIR 5000 System spectrophotometer (Silver Spring, MD, USA) equipped with a transport module. Round sample cups (3.8 cm diameter x 1cm thick quartz windows) were filled with each sample (about 5 g) so that the surface and thickness remained uniform during spectral collection. The instrument measures diffuse reflectance and automatically converts it into relative absorbance (log 1/R) to obtain a linear correlation with the concentration of the product's chemical constituents (Martens, Nielsen, & Engelsen, 2003). Thirty-two successive scans with 700 points (wavelengths) from each sample were collected within a wavelength range from 1100 nm to 2500 nm at 2-nm intervals.

2.5 Chemometric analysis

An analysis of variance (ANOVA) was used to determine the differences in pH and extrinsic color among samples. Data were statistically processed using Statgraphics Centurion XVI (Manugistics Inc., Rockville, MD, USA). Simultaneously, color parameters C*, h* and L* and pH were used in a principal component analysis (PCA) to show the samples and their relationship. Before the analysis, an autoscaling was performed in order to improve the weights of the variable with small values.

A multivariate analysis was conducted by a qualitative analysis and a quantitative analysis by The Unscrambler v10.4 (CAMO Software AS, OSLO, Norway). For the qualitative analysis, a PCA and a partial least squares discriminant analysis (PLS-DA) was performed. The PCA was run with raw data, while the PLS-DA (Berrueta, Alonso, & Héberger, 2007) was constructed after applying spectra 2nd derivative Savitzky-Golay smoothing (2nd derivative S-G) (Savitzky & Golay, 1951) and orthogonal signal correction (OSC). Both pre-treatments were applied to acquire useful information, improve the signal-to-noise ratio and remove systematic variation from the predictor matrix X unrelated, or orthogonal, to matrix Y (Wold, Antti, Lindgren, & Öhman, 1998; Pizarro *et al.*, 2004). For the quantitative analysis,

116

a partial least squares (PLS) regression analysis was performed. In order to evaluate and correct the multiplicative and additive effects caused by different light scattering in the spectroscopic measurement (Cozzolino *et al*., 2011; Stohner *et al*., 2012), four PLS models were tested. The PLS were constructed using the raw spectrum and by applying three pre-treatments to the spectrum: 2nd derivative S-G, OSC and the combination of them.

2.5.1 Developing calibration models

Two databases were used for the analysis. The first database consisted of 468 spectra and 700 variables (wavelengths, nm) was used for the PCA and PLS models. For the PLS-DA classification samples were divided into three categories (0=Cocoa; 1=Adulterated samples and 2=Carob flour) and a second database with 142 spectra and 700 variables was created to balance the number of samples that belonged to each category. Moreover, the spectra of each database were randomly separated into two different data sets. A set with 70% of the spectra was used to create and evaluate the model by leave-one-out cross-validation. The other set, with 30% of the remaining samples, was used for external validation. The relative performance of the constructed models was assessed by the required number of latent variables (LVs), the coefficient of determination for calibration (R^2_C), the root mean square error of calibration (RMSEC), the coefficient of determination for cross validation (R^2_{CV}) and the root mean square error of leave-one-out cross validation (RMSECV). A model can be considered good when a few LVs are required, and when it has low RMSEC and RMSECV and high R^2_C and R^2_{CV}. A cut-off value of $\pm$ 0.5 was used for the classification of the samples (Dong, Zhao, Hu, Dong, & Tan, 2017).

2.5.2 External validation

To assess the models' predictive capability, the coefficient of determination for prediction (R^2P), the root mean square error of prediction (RMSEP), the ratio of

117

prediction deviation (RPD = SD/RMSEP), where SD was the standard deviation of the Y-variable in the prediction set, and bias were used. The RPD is more meaningful than only looking at the error of prediction. An RPD value lower than 2 is considered insufficient for application, one between 2 and 2.5 is considered for approximate quantification, and values between 2.5 and 3 are taken as a good model, while models with RPD values above 3 can be considered excellent and most reliable for analytical tasks (Sunoj, Igathinathane & Visvanathan, 2016). The bias estimates the difference between the experimental value and NIR predictions, and can be positive or negative. Positive values indicate that the model overestimates, while negative values suggest otherwise. Higher bias values indicate that NIR predictions vary significantly from the experimental values (Cantor, Hoag, Ellison, Khan, & Lyon, 2011), so it is better if it comes close to zero.

3. Results and discussion

3.1 Raw materials characterization

Table 1 contains the color parameters and pH values of the different raw materials. As observed, the obtained pH values ranged from 5.3 (NC1) to 7.9 (SAC3). According to these values and following the Miller Classification, twelve samples were considered natural cocoas (NC; 5<pH<6), three samples light alkalized cocoas (LAC; 5<pH<6.2), tree samples medium alkalized cocoas (MAC; 7.2<pH<7.6) and three samples strong alkalized cocoas (SAC; PH > 7.6). pH can be used as an indicator of the degree of alkalization that occurs during production because the pH value of cocoa powder is related to the amount and type of alkali used in the process (OLAM Cocoa Manual, 2017; Pérez, Lerma, Fuentes, Palomares, & Barat, 2016). The inclusion of cocoas with different degrees of alkalization during the model-building phase assures that it might be used with independence of the cocoa powder processing.

The lightness (L*) values measured in the cocoa samples ranged from 31 (SAC1) to 50 (NC3). The maximum lightness value appeared in a NC sample (NC3). The L* value progressively lowered according to the degree of alkalization to the minimum value in the SAC samples with a very dark color. The differences in the lightness in the NC samples (NC1, NC2 and NC3) could be due to a different geographical origin or to distinct processing in the fermentation or roasting stages (Afoakwa, *et al*., 2014).

The chroma (C*) values oscillated between 11 (SAC1) and 22 (NC2). As seen in Table 1, the higher the alkalization degree, the lower purity becomes.

Hue (h*), unlike the other parameters, does not follow a linear relationship with an increased pH value. Cocoa samples evolve from a more yellow-orange hue (h * = 60) to a more orange-red one (h* = 43) in the alkaline cocoa samples.

The pH of carob flours ranged from 4.5 to 5.1, with no trend observed between the pH value and the degree of toasting samples. Thus, carob samples could be added to the NC beans in high proportions without significantly changing the mixture's pH value.

The L* values in the carob flours ranged from 34 (DCF) to 49 (LCF), which meant that lightness progressively lowered as the degree of roasting increased. When these values were compared with those of cocoa, were found no statistical differences (p<0.05) between the NC samples and natural carob meal (LCF samples), nor between the lightness of MAC and SAC samples and roasted carob (MCF and DCF). These minor differences in lightness would favor the adulteration of cocoa with carob meal.

The chroma (C*) of the samples also decreased as the degree of roasting rose, with values of 23.7 for LCF samples that lowered to 13 for strong roasted carob (DCF samples). When comparing the C* values between cocoa and carob, we found a similarity between both. Thus the C* values would be the equivalent between natural cocoa and natural carob meal, and between medium/strong cocoa beans and roasted carob.

The hue (h*) values for the carob flours gave no significant differences with an increasing degree of roasting, but only a slight decrease. The values obtained for the carob flour samples were 61 on average. These values coincided with those observed in the NC samples.

Cocoa color parameters are generally affected by several factors, including the degree of roasting and alkalization. The strong alkalized ones were dark, while the natural ones were lighter. The roasting result was darkened cocoa or carob due to the formation of brown pigments (Zyzelewicz, Krysiak, Nebesny, & Budryn, 2014), with changes noted in the values of the individual color parameters.

Table 1. Color parameters and pH (mean and standard deviation) values for the carob and cocoa pure samples.

| Product	Color Parameters			pH ± sd
	L* ± sd	C* ± sd	h* ± sd	
LCF1	48.6±0.4[de]	23.6±0.3[e]	61.0±0.3[b]	5.03±0.01[a]
LCF2	47.70±0.06[de]	24.1±0.2[e]	60.98±0.11[b]	5.12±0.01[a]
LCF3	46.1±0.2[de]	26.3±0.7[e]	61.5±0.3[b]	4.67±0.01[a]
LCF4	44.17±0.3[de]	20.7±0.2[e]	61.1±0.3[b]	4.911±0.01[a]
MCF	37.6± 0.4[ab]	16.9±0.7[bc]	60.2±0.5[a]	4.851±0.01[a]
DCF	34.5±1.5[a]	12.9±0.9[a]	60±2[a]	4.817±0.01[a]
NC1	48.7±0.2[e]	20.1±0.5[de]	58.8±0.4[c]	5.391±0.01[a]
NC2	48.33±0.13[e]	22.3±0.4[de]	59.5±0.3[c]	5.46±0.01[b]
NC3	50.3±0.6[e]	22.19±1.02[de]	60.0±0.4[c]	5.70±0.01[b]
LAC1	42.3±0.6[c]	22.4±0.7[cd]	54.3±0.4[c]	6.901±0.01[c]
LAC2	44.2±0.5[c]	18.63±1.02[cd]	55.0±0.9[c]	6.96±0.02[c]
LAC3	41.7±0.5[b]	19.80±0.13[bc]	54.5±0.5[c]	6.98±0.01[d]
MAC1	44.9±1.5[c]	18±2[cd]	55.7±0.6[c]	7.24±0.01[c]
MAC2	41.9±0.7[b]	18.0±0.6[bc]	54.2±0.5[c]	7.34±0.03[d]
MAC3	35.85±1.05[b]	16.0±0.8[bc]	43.0±0.6[c]	7.43±0.01[d]
SAC1	32.1±0.8[a]	11.6±0.9[b]	46.5±0.6[c]	7.81±0.01[e]
SAC2	39.4±0.5[a]	19.76±0.99[b]	51.4±0.8[c]	7.84±0.01[e]
SAC3	40.1±0.2[a]	17.3±0.8[b]	53.2±0.6[c]	7.92±0.01[e]

Values in the same column followed by the same letter(s) are not significantly different according to ANOVA at a 95% Confidence level. For cocoas (N): Natural cocoa (NC), light alkalized cocoa (LAC), medium alkalized cocoa (MAC) and strong alkalized cocoa (SAC). For carob flours (A): light carob flour (LCF), medium carob flour (MCF) and dark carob flour (DCF).

In order to know how the physico-chemical properties explained the different characteristics between the cocoa and carob flour samples, a PCA was performed

with the pH and color parameters. Figure 1 shows the two-dimensional scatter plot of scores for two principal components (PCs) from projection results and allows the visualization of the distribution of the scores of the samples of cocoa and carob powder. The two PCs explain over 94% of the variation. The first PC explains 78 % of the variance and is related with the roasted degree of the carob, or alkalization of the cocoa powder, whereas the second PC explains 16 % of the total variance and is related with the difference between level of alkalized cocoa and the carob flour.

The natural cocoa NC and light carob flour LCF scores were close, which indicated that these samples were related and had similar pH characteristics and color parameters. The positive scores on component 1 and component 2 corresponded to the samples with different degrees of alkalization. This position and the loading values of the variables led to the conclusion that the samples with low lightness and high pH were the alkaline cacao samples, while the samples with low lightness and low pH were those of roasted carob flour (Dark (DCF) and medium (MCF)). This agrees with the results presented by other authors (Bulca, 2016; Yousif & Alghzawi, 2000), which indicated that carob flour could not be visually separated from cocoa powder, not even when the other groups of the alkalized and roasting samples were blended.
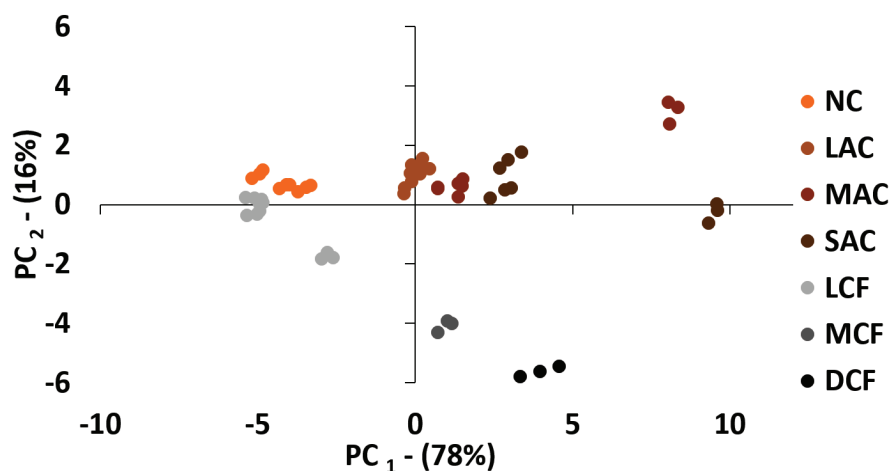
Figure 1. Score plot of the first and the second principal components of the PCA model using color parameters L*, C*, h* and pH of pure carob and cocoa powder samples (n=18, in triplicate). NC: natural Cocoa; LAC: light alkalized cocoa; MAC: medium alkalized cocoa; SAC: strong alkalized cocoa; LCF: light carob flour; MCF: medium carob flour; DCF: dark carob flour.

3.2 Spectral differences analysis of carob and cocoa powder

The spectra of the relative absorbance of cocoa powder and carob flour are represented in **Figure 2** (a, b). All the cocoa spectra display a similar absorbance pattern, this pattern differs between cocoa and carob flour in relation to the absorbance intensity.
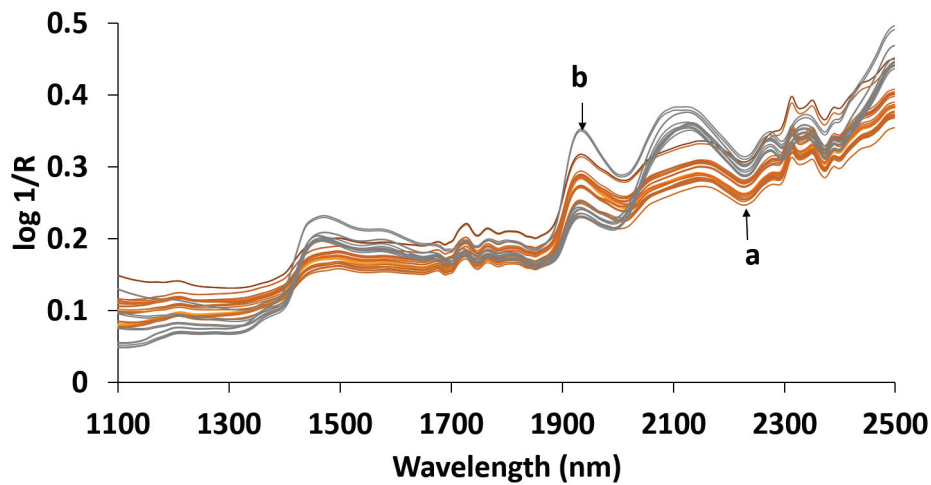
Figure 2. Spectra with raw data within the 1100 - 2500 nm range (a) Cocoa. (b) Carob flour.

Raw data were preprocessed by applying the 2nd derivative S-G and OSC. Examples of the pretreated spectra of cocoa (brown) and carob (gray) are shown in **Figure 3**. As observed, after this pretreatment the differences between both spectra types became more evident than in the untreated spectra. It can be stated how divergence pointed between both spectra types being located especially in the magnitude of reflectance at 1438, 1728, 2312, 2324, 2350 nm. As expected from the compositional differences, between cocoa powder and carob flours, these wavelengths were associated with the vibration of the functional groups that cocoa powder contains like theobromine and caffeine (1728 nm) (Cozzolino *et al.*, 2011), and epicathechin (2312, 2324 nm) (Esteban, González, & Pizarro, 2004; Teye & Huang, 2015).
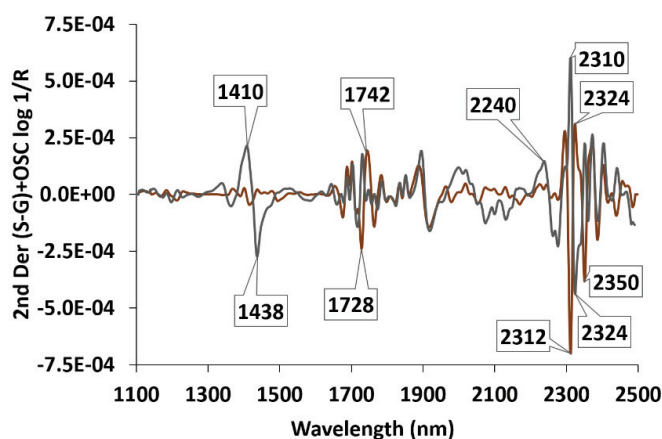
Figure 3. Second derivative, Savitzky Golay smoothing and orthogonal signal correction pretreated the cocoa (brown) and carob (grey) spectra within the 1100 - 2500 nm range.

3.2 Classification model

A PCA was performed as a non-supervised learning algorithm with the raw spectra data to evaluate the relationship among samples. **Figure 4** shows the score plot of the first two principal components (PC). The first two PC explain 91% of the total variance among the samples. The first PC explains 71% of total variance and might be related to sample processing. The different natural cocoas are found in the negative region, whereas the alkalinized samples are distributed across the negative and positive regions. These differences could be due to alkali, the stage in which it has been alkalized (bean or cake), and the degree of alkalization that can produce different color changes (red or dark brown) (Miller *et al*., 2008). The second PC explains 20% of variability and might be related to the percentage of cocoa powder in the sample. Pure cocoa powders are located in the positive region, while pure carob flours are found in the negative region. The samples with different levels of adulteration lie in the middle: low (0-20%), medium (20-40%) and high (40-60%).

The wavelengths that corresponded to the highest loading values were 1100, 1464, 1936, 2108, 2276, 2330 and 2486 nm for the first PC, and 1116, 1324 1460, 1576, 1728, 1914, 1976, 2106, 2262, 2310 and 2494 nm for the second PC. The

125

wavelengths from 971 and 1400 nm were related to the ascending part of the water first overtone absorption peak O–H stretching bonds at 1722 nm C-H stretching was also present, which are associated with water and sugar content (Álvarez *et al.*, 2012; Cozzolino, Smyth, & Gishen, 2003; X. Y. Huang *et al.*, 2014; Talens *et al.*, 2013). The wavelengths at 1736 and 2319-2328 nm were related to the absorption of the C–H bonds, CH3 combination and C-C stretching. These are features of fatty acids, proteins and polysaccharides in cocoa powder and could be associated with a fat content of approximately 10-12% (Veselá *et al.*, 2007; Westad, Schmidt, & Kermit, 2008). The absorption bands of 1728, 2108 and 2494 nm coincided approximately with those that were used to predict the total fat content in cocoa beans by (Ribeiro, Ferreira, & Salva, 2011; Teye & Huang, 2015). Variations were related to the compositional characteristics of the cocoa categories and the adulterant carob powder. The found wavelengths were similar to a study performed in cocoa beans (Teye *et al.*, 2015b). Therefore, absorption in wavelengths (as a result of vibrational reactions) contains chemical information that helps explain the observed differences between the carob and cocoa powder pure samples and their several adulteration proportions.

Since the generated spectra correspond to an adulteration level on a continuous scale, it was not possible to see well-separated groups (high, medium and low adulterated) in this PCA, especially for the percentages that fell within the limits. With this information, a PLS-DA analysis was created to generate a model with categorized spectra, which allowed the detection of gross adulterations levels.
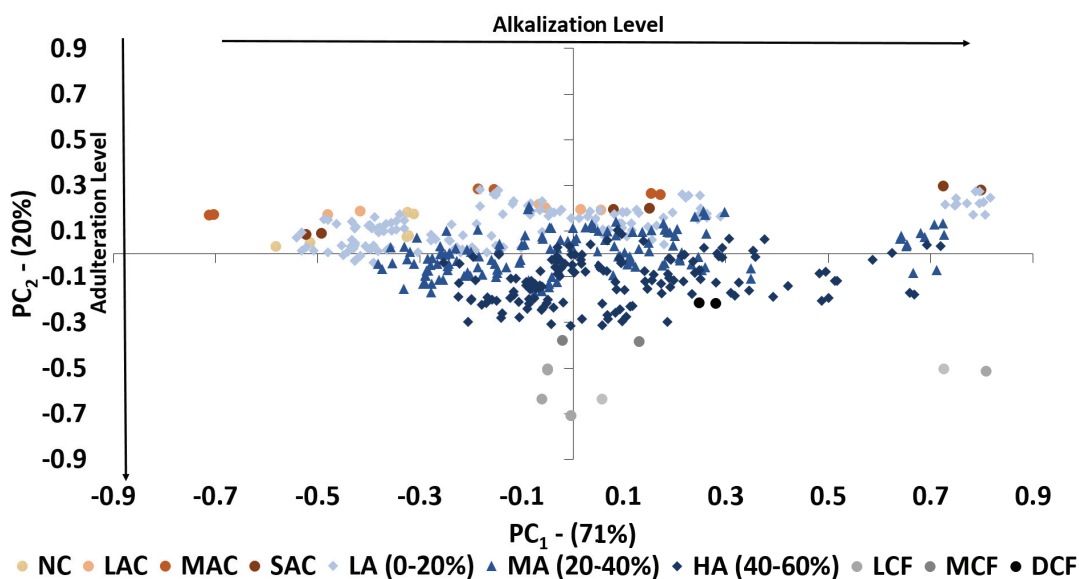
Figure 4. The NIR PCA score plot to separate pure cocoa powder and different levels of adulteration with carob flour (high adulteration HA (40-60%), low adulteration LA (0-20%) and medium adulteration MA (20-40%)).

As the PCA was unable to see samples in the different groups according to their adulteration percentages, a qualitative model that used the supervised PLS-DA was employed. In order to improve the model's accuracy, the original spectra were pre-processed using 2nd derivative S-G (9-point window, second-order polynomial) and an OSC. For the PLS-DA (**Figure 5**), three latent variables (LVs) were generated with most of the variation (67%) explained by the first LV and 12% by the second. In this way, separation was achieved mainly by using the first latent variable with the most negative scores for the pure cocoa samples, and the most positive scores related to the adulterate samples and carob powder (pure adulterant). In visually terms, the scores plot differences among the 100% cocoa powder, adulterated cocoa powders and 100% carob powder indicated the possibility of using this approach to quickly screen for adulteration. The determination coefficient (R^2) of this PLS-DA model was 0.969. The cross-validation determination coefficient (R^2_{CV}), based on full cross-validation, was 0.901. Those values indicate the goodness of the classification model.

127
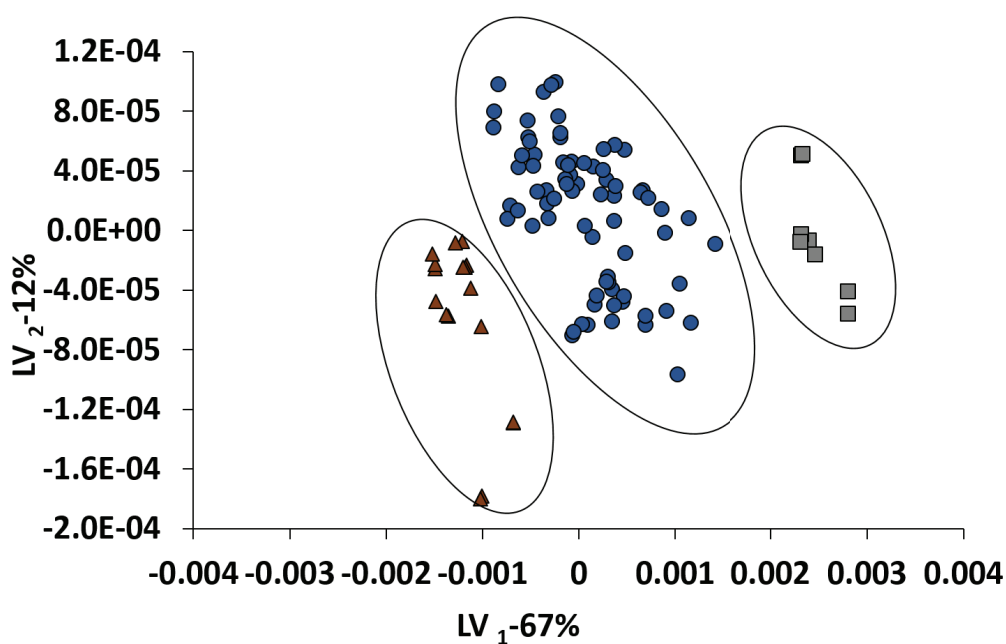
Figure 5. NIR PLS-DA score plot from latent variable 1 and 2, pure cocoa (brown triangles) carob powder (gray squares) and adulterations (blue circles).

In order to measure the robustness of the PLS-DA model, validation with an external data set was performed. Table 2 shows the model's capability to classify 100% of the samples in their corresponding groups (cocoa, carob or adulterated samples).

Table 2. Results for classification accuracy of the PLS-DA model.

	Cocoa	Carob	Adulterated	Classification
Cocoa	7	0	0	100%
Carob	0	4	0	100%
Adulterated	0	0	32	100%

3.3 Adulterant Prediction

A PLS was performed with the calibration set. The prediction was done with the validation set. The models were constructed by applying different pre-treatments to

the studying the spectra. The statistical indicators of the goodness of fit of each model are presented in **Table 3**.

Table 3. Results of the PLS models constructed for the prediction of carob flour content in cocoa powders.

Pre-treatment	#LV	Calibration		Cross-validation		Prediction			
		R^2_C	RMSEC	R^2_{CV}	RMSECV	R^2_P	RMSEP	Bias	RPD
Raw data	7	0.951	4.530	0.945	4.785	0.961	4.397	0.197	4.65
2nd Der. S-G	5	0.978	3.082	0.974	3.28	0.979	3.271	0.749	6.25
OSC	1	0.975	3.165	0.975	3.214	0.974	3.555	0.474	5.75
2nd Der. S-G + OSC	*1*	*0.980*	*2.856*	*0.979*	*2.897*	*0.974*	*3.237*	*0.626*	*6.32*

2nd Der. S-G = Second derivative-Savitzky Golay; OSC = Orthogonal signal correction; #LV = latent variables; R^2_C = coefficient of determination for calibration; RMSEC = root mean square error of calibration; R^2_{CV} = coefficient of determination for cross-validation; RMSECV = root mean square error of cross-validation; R^2_P = coefficient of determination for prediction; RMSEP = root mean square error of prediction; Bias = estimation of the difference between the experimental value and NIR predictions; RPD = ratio of prediction deviation.

Good models were obtained with high R^2 values and low RMSE values for the calibration, cross-validation and prediction, depending on spectral data processing. The RPD values were higher than 3, which meant that all these models, even the model without the preprocessing data, could be considered excellent and most reliable for the analytical tasks. This indicated that the multiplicative and additive effects in the spectra of cocoa powder, and with the equipment used for the

measurements in this study, were minimal due to even the model without pretreatment to correct that effects was excellent. However, it is important to point out that the models with pre-treatments obtained a smaller number of LV, which made the model more parsimonious. Figure 6 presents the observed (x axis) versus predicted (y axis) values. The predicted values were obtained with a model that used 2nd Derivative S-G and OSC. We can observe that the PLS algorithm predicted very well with an R^2_{CV} of 0.979 and an RMSECV of 2.897%. The prediction of the external validation group gave a low RMSEP of 3.237% and an R^2_P of 0.974. The similarity among RMSEC, RMSECV and RMSEP indicated that the possibility of over-fitting the model was very low and confirmed its good prediction capacity. The 2nd derivative S-G and the OSC pretreatment improved the RPD, which was 35.48% higher compared to the PLS model with the raw data, and used only one latent variable (LV). Other studies have found good models with one LV when orthogonal signal correction was used (Esteban *et al.*, 2004). The relative notorious improvement of the RPD in the pretreated model could be due to the NIR signal being affected by environmental (moisture) and physical factors (product's particle size distribution). According to Huang *et al.*, those factors generated light scattering and, consequently, significant differences arose. Additionally, these factors affect the effective sample pathlength and result in additive, multiplicative and wavelength-dependent effects. In some cases, wavelength-dependent scattering is related with baseline shifts, tilt or curvature scaling variation. In certain instances, spectra variations mask any subtle chemical variation, which can produce inaccurate results. Thus pretreatment is effective for cushioning the aforementioned effects (Huang *et al.*, 2010).
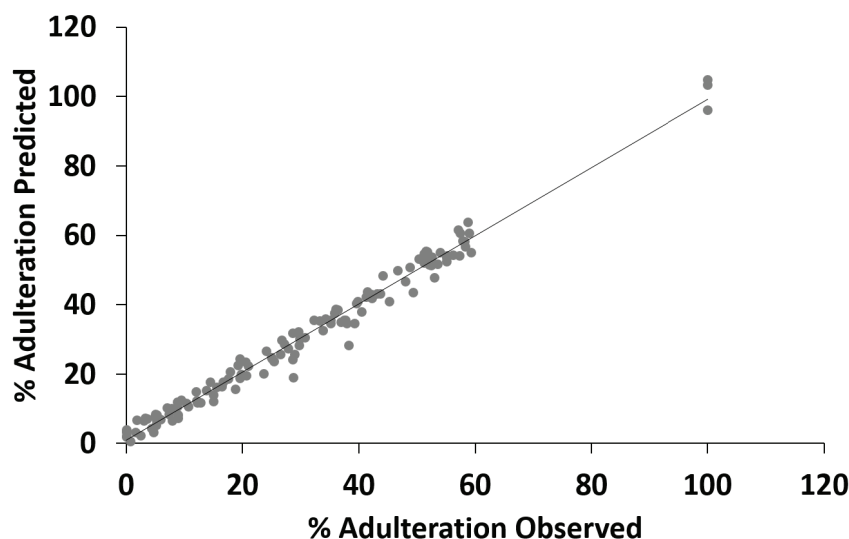
Figure 6. Predicted versus observed values of the adulterant percentages (n=140) of the pure cocoa and carob powder at different levels of adulterated samples.

4. Conclusions

Near infrared spectroscopy (NIR) combined with PLS-DA and PLS statistical models has been shown to be a rapid effective method to identify adulterations of cocoa powder with Carob flour, regardless of the alkalization or roasting level. In contrast, these adulterations would not be readily detectable by routine techniques such as determination of pH analysis and color measurement.

With the PLS-DA analysis, all (100%) the samples were correctly classified into three groups: cocoa, carob flour and mixtures. The PLS analysis enabled the percentage of adulteration to be calculated with the samples. The PLS model was obtained with one factor with an R^2 of 0.979 and 0.974, and a mean squared error of 2.9 and 3.2 for the calibration and external validation sets, respectively.

This technique is, therefore, an important tool for cocoa merchants, who will be able to better control the product's quality by avoiding the use of destructive techniques that require complex sample preparations or techniques that imply much

expense for companies. Given the excellent results obtained, we expect this method to become increasingly important in the cocoa industry and to reduce food fraud.

References

Afoakwa, E. O., Budu, A. S., Mensah-brown, H., Felix, J., & Ofosu-ansah, E. (2014). Effect of Roasting Conditions on the Browning Index and Appearance Properties of Pulp Pre-Conditioned and Fermented Cocoa (Theobroma Cacao) Beans. *J Nutrition Health Food Sci, 2*(1), 1–5.

Álvarez, C., Pérez, E., Cros, E., Lares, M., Assemat, S., Boulanger, R., & Davrieux, F. (2012). The use of near infrared spectroscopy to determine the fat, caffeine, theobromine and (-)-epicatechin contents in unfermented and sun-dried beans of Criollo cocoa. *Journal of Near Infrared Spectroscopy, 20*(2), 307.

Arrighi, W. J., Hartman, T. G., & Ho, C. T. (1997). Carob bean aroma dependence on roasting conditions. *Perfumer & flavorist, 22*(1), 31-41.

Berrueta, L. A., Alonso, R. M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A, 1158*(1–2), 196–214.

Bonvehí, J. S. (2005). Investigation of aromatic compounds in roasted cocoa powder. *European Food Research and Technology, 221*(1–2), 19–29.

Bulca, S. (2016). Some properties of carob pod and its use. Scientific Bulletin. Series F. *Biotechnologies, Vol. XX*, 142–147.

132

Cantalejo, M. J. (1997). Effects of Roasting Temperature on the Aroma Components of Carob (Ceratonia siliqua L.). *Journal of Agricultural and Food Chemistry, 45*(4), 1345–1350.

Cantor, S.L., Hoag, S.W., Ellison, C.D., Khan, M.A., Lyon, R.C., 2011. NIR spectroscopy applications in the development of a compacted multiparticulate system for modified release. *AAPS PharmSciTech 12*, 262–278.

Capuano, E., Boerrigter-Eenling, R., Koot, A., & van Ruth, S. M. (2015). Targeted and Untargeted Detection of Skim Milk Powder Adulteration by Near-Infrared Spectroscopy. *Food Analytical Methods, 8*(8), 2125–2134.

Cozzolino, D., Smyth, H. E., & Gishen, M. (2003). Feasibility Study on the Use of Visible and Near-Infrared Spectroscopy Together with Chemometrics to Discriminate between Commercial White Wines of Different Varietal Origins. *Journal of Agricultural and Food Chemistry, 51*(26), 7703–7708.

Cozzolino, Cynkar, W. U., Shah, N., & Smith, P. (2011). Multivariate data analysis applied to spectroscopy: Potential application to juice and fruit quality. *Food Research International, 44*(7), 1888–1896.

Dong, W., Zhao, J., Hu, R., Dong, Y., & Tan, L. (2017). Differentiation of Chinese robusta coffees according to species, using a combined electronic nose and tongue, with the aid of chemometrics. *Food Chemistry, 229*, 743–751.

Ellis, D. I., Brewster, V. L., Dunn, W. B., Allwood, J. W., Golovanov, A. P., & Goodacre, R. (2012). Fingerprinting food: current technologies for the detection of food adulteration and contamination. *Chemical Society Reviews, 41*(17), 5706.

Esteban, I., González, J. M., & Pizarro, C. (2004). Prediction of sensory properties of espresso from roasted coffee samples by near-infrared spectroscopy. *Analytica Chimica Acta, 525*(2), 171–182.

Fadel, H. H., Mageed, M. A. A., Samad, A. K. M. A., & Lotfy, S. N. (2006). Cocoa substitute: Evaluation of sensory qualities and flavour stability. *European Food Research and Technology, 223*(1), 125-131.

Haughey, S. A., Galvin-King, P., Ho, Y.-C., Bell, S. E. J., & Elliott, C. T. (2015). The feasibility of using near infrared and Raman spectroscopic techniques

to detect fraudulent adulteration of chili powders with Sudan dye. *Food Control, 48*, 75–83.

Helton, J. C., & Davis, F. J. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering and System Safety, 81*(1), 23–69.

ICCO (2017). International Cocoa Organization Dailyy Prices of cocoa Beans https://www.icco.org/statistics/cocoa-prices/daily-prices.html, Consulted: 13th January 2018.

Krysiak, W. (2006). Influence of roasting conditions on coloration of roasted cocoa beans. *Journal of Food Engineering, 77*(3), 449-453.

Huang, J., Romero, S., & Moshgbar, M. (2010). Practical Considerations in Data Pre-treatment for NIR and Raman Spectroscopy. American Pharmaceutical Review, 13(6). http://www.americanpharmaceuticalreview.com. Consulted: 2 December 2017

Huang, X. Y., Teye, E., Sam-Amoah, L. K., Han, F. K., Yao, L. Y., & Tchabo, W. (2014). Rapid measurement of total polyphenols content in cocoa beans by data fusion of NIR spectroscopy and electronic tongue. *Analytical Methods, 6*(14), 5008–5015.

Li, Y., Feng, Y., Zhu, S., Luo, C., Ma, J., & Zhong, F. (2012). The effect of alkalization on the bioactive and flavor related components in commercial cocoa powder. *Journal of Food Composition and Analysis, 25*(1), 17-23.

Li, X., Zhang, Y., & He, Y. (2016). Rapid detection of talcum powder in tea using FT-IR spectroscopy coupled with chemometrics. Scientific Reports, 6(February), 30313.

Lohumi, S., Lee, S., Lee, H., & Cho, B. K. (2015). A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. *Trends in Food Science and Technology, 46*(1), 85–98.

Martens, H., Nielsen, J. P., & Engelsen, S. B. (2003). Light scattering and light absorbance separated by extended multiplicative signal correction. Application to

near-infrared transmission analysis of powder mixtures. *Analytical Chemistry, 75*(3), 394–404.

Miller, K. B., Hurst, W. J., Payne, M. J., Stuart, D. A., Apgar, J., Sweigart, D. S., & Ou, B. (2008). Impact of alkalization on the antioxidant and flavanol content of commercial cocoa powders. *Journal of Agricultural and Food Chemistry, 56*(18), 8527–8533.

Moore, J. C., Spink, J., & Lipp, M. (2012). Development and Application of a Database of Food Ingredient Fraud and Economically Motivated Adulteration from 1980 to 2010. *Journal of Food Science, 77*(4).

OLAM (2017). De Zaan® Cocoa Manual. The Netherlands: Archer Daniels Midland Company BV.

Pérez, É., Lerma, M. J., Fuentes, A., Palomares, C., & Barat, J. M. (2016). Control of undeclared flavoring of cocoa powders by the determination of vanillin and ethyl vanillin by HPLC. *Food Control, 67*, 171–176.

Pizarro, C., Esteban, I., Nistal, A.-J., & González, J.-M. (2004). Influence of data pre-processing on the quantitative determination of the ash content and lipids in roasted coffee by near infrared spectroscopy. *Analytica Chimica Acta, 509*, 217–227.

Ribeiro, J. S., Ferreira, M. M. C., & Salva, T. J. G. (2011). Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy. *Talanta, 83*(5), 1352–1358.

Rosa, C. S., Tessele, K., Prestes, R. C., Silveira, M., & Franco, F. (2015). Effect of substituting of cocoa powder for carob flour in cakes made with soy and banana flours. *International Food Research Journal, 22*(5), 2111–2118.

Salem, E. M., & Ohaad Fahad, A. A. (2012). Substituting of Cacao by Carob Pod Powder In Milk Chocolate Manufacturing. *Australian Journal of Basic and Applied Sciences, 6*(3), 572–578.

Savitzky, A., & Golay, M. (1951). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Z. Physiol. Chem. *Chem. & Ind, 40*(42).

Article 2

Shankar, M. U., Levitan, C. A., Prescott, J., & Spence, C. (2009). The influence of color and label information on flavor perception. *Chemosensory Perception, 2*(2), 53–58.

Stohner, J., Lukas, B., Suter, M., Zucchetti, B., Deuber, F., & Hobi, F. (2012). NIRS of chocolate and its chemometric analysis. *Newfood, 15*(6), 21–28.

Sunoj, S., Igathinathane, C., & Visvanathan, R. (2016). Nondestructive determination of cocoa bean quality using FT-NIR spectroscopy. *Computers and Electronics in Agriculture, 124*, 234–242.

Talens, P., Mora, L., Morsy, N., Barbin, D. F., ElMasry, G., & Sun, D.-W. (2013). Prediction of water and protein contents and quality classification of Spanish cooked ham using NIR hyperspectral imaging. *Journal of Food Engineering, 117*(3), 272–280.

Teye, E., & Huang, X. (2015). Novel Prediction of Total Fat Content in Cocoa Beans by FT-NIR Spectroscopy Based on Effective Spectral Selection Multivariate Regression. *Food Analytical Methods, 8*(4), 945–953.

Teye, E., Huang, X., Sam-Amoah, L. K., Takrama, J., Boison, D., Botchway, F., & Kumi, F. (2015). Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis. *Food Chemistry, 176*, 403–410.

Veselá, A., Barros, A. S., Synytsya, A., Delgadillo, I., Čopíková, J., & Coimbra, M. A. (2007). Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder. *Analytica Chimica Acta, 601*(1), 77–86.

Westad, F., Schmidt, A., & Kermit, M. (2008). Incorporating chemical band-assignment in near infrared spectroscopy regression models. *Journal of Near Infrared Spectroscopy, 16*(1), 265.

Wold, S., Antti, H., Lindgren, F., & Öhman, J. (1998). Orthogonal signal correction of near-infrared spectra. Chemometrics and Intelligent, 44, 175–185.

Yousif, A. K., & Alghzawi, H. M. (2000). Processing and characterization of carob powder. *Food Chemistry, 69*(3), 283–287.

Zyzelewicz, D., Krysiak, W., Nebesny, E., & Budryn, G. (2014). Application of various methods for determination of the color of cocoa beans roasted under variable process parameters. *European Food Research and Technology, 238*(4), 549–563.

3.2 CHAPTER 2. Prediction of the content of bioactive compounds of cocoa powders.

3.2.1 Changes in methylxanthines and flavanols during cocoa powder processing and their quantification by near-infrared spectroscopy

Maribel Alexandra Quelal-Vásconez,[a] María Jesús Lerma-García,[a] Édgar Pérez-Esteve,[a] Alberto Arnau-Bonachera,[b,c] José Manuel Barat,[a] Pau Talens,[a*]

[a]Departamento de Tecnología de Alimentos, Universitat Politècnica de València. Camino de Vera s/n, 46022, Spain

[b]Institute for Animal Science and Technology, Universitat Politècnica de València - Universidad de Valencia. Departamento de Química Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain

[c]Biomedical Research Institute (PASAPTA-Pathology group), Veterinary School, Universidad Cardenal Herrera-CEU, CEU Universities, Av. Seminario s/n, 46113 Moncada, Valencia, Spain

Abstract

Variation in methylxanthines (theobromine and caffeine) and flavanols (catechin and epicatechin) was studied in a large set of cocoa powders (covering different origins, processing parameters and alkalisation levels). The content of these compounds was established by high-performance liquid chromatography (HPLC), whose results showed that the alkalisation process lowered the content of all analytes, whose loss was more evident in flavanols. Therefore, the determination of these analytes in a huge set of samples allowed not only better knowledge of the concentration variability in natural commercial cocoas from different origins, but also the understanding of the effect that industrial alkalisation has on these contents. The feasibility of reflectance near-infrared spectroscopy (NIRS) combined with partial least square (PLS) to non-destructively predict these contents, was also evaluated. All the analytes were generally well predicted, with predictions for methylxanthines (R2P 0.819–0.813 and RMSEP 0.068–0.022%, and bias 0.005 and 0.007 for theobromine and caffeine, respectively) and for flavanols (R2P 0.830–0.824; RMSEP 8.160–7.430% and bias −1.440 and −1.034 for catechin and epicatechin, respectively). Thus NIRS could be an alternative fast reliable method for the routine assessment of these analytes in the cocoa industry.

Keywords: cocoa powder, methylxanthines, flavanols, HPLC, near-infrared spectroscopy.

Abbreviations: HPLC, high performance liquid chromatography; GC, gas chromatography; IR, infrared spectroscopy; NIRS, near infrared spectroscopy; PLS, partial least square; PCA, principal component analysis; LV, latent variable; RMSEC, root mean square error of calibration; RMSECV, root mean square error of cross-validation; RMSEP, root mean square error of prediction; R^2_C, coefficient of determination for calibration; R^2_{CV}, coefficient of determination for cross-validation; R^2_P, coefficient of determination for prediction; RPD, ratio of prediction deviation;

LOD, limit of detection; LOQ, limit of quantification; S-G, Savitzky-Golay; OSC, orthogonal signal correction.

1. Introduction

Cocoa powder is the most important raw material of confectionery products, chocolate-flavoured bakeries, ice-creams and drinks (Miller *et al*., 2008). Apart from technologic properties, cocoa (*Theobroma cacao* L.) and its derived products are well considered for being a rich source of methylxanthines and polyphenols (Langer, Marshall, Day, & Morgan, 2011).

Methylxanthines (i.e. theobromine and caffeine) are pharmacologically active alkaloids responsible for a bitter cocoa taste and desirable physiological effects; e.g. stimulation of the central nervous system and gastric secretion, diuresis, bronchodilation, and stimulation of skeletal muscles in high doses (Franco, Oñatibia-Astibia, & Martínez-Pinilla, 2013). They also display health benefits in diseases involving cell death in the nervous system (Oñatibia-Astibia, Franco, & Martínez-Pinilla, 2017).

The main type of polyphenols (known for their demonstrated antioxidant and anti-inflammatory properties) in cocoas are flavanols. This family of compounds includes catechin and epicatechin (monomeric species) and procyanidins (oligomeric and polymeric fractions). Among them, epicatechin is the most abundant flavanol in cocoa and accounts for 35% of the total polyphenolic fraction (Lacueva *et al*., 2008).

While producing cocoa powder from cocoa beans, seeds are primarily fermented, dried and roasted. Then broken beans (nibs) are ground, heated and liquefied. The product of these operations, cocoa liquour, is pressed to obtain two different fractions: cocoa powder and butter. Optionally, nibs or cake can be treated with an alkali dissolved in water. This alkalisation reduces acidity, bitterness and astringency by improving and enhancing the aromatic features of cocoa powder (Kongor *et al*., 2016). Alkalisation also allows colour development by transforming

the typical light brown hues of natural cocoa powder into reddish or very dark ones. These colour changes derive from a sequence of chemical reactions between alkalizing agents and pigments in the presence of water and oxygen at high temperatures. Finally, the dispersibility of cocoa powder increases with alkalisation. This property facilitates the use of cocoa powders in different industries like dairy products (Afoakwa, Paterson, Fowler, & Ryan, 2008).

Cocoa powder processing not only strongly affects the sensory properties of cocoa and derived products, but also alters flavonoid and methylxanthine fractions (Payne, Hurst, Miller, Rank, & Stuart, 2010). However, studies to date have used small sample numbers and have, thus provided conclusions that can be read only as behaviour tendencies.

The commonest techniques employed to analyse methylxanthines and flavanols from cocoa extracts or isolated fractions are high-performance liquid chromatography (HPLC) or gas chromatography (GC) (Cádiz-Gurrea *et al*., 2014; Elwers, Zambrano, Rohsius, & Lieberei, 2009; Fayeulle *et al*., 2018, Humston, Knowles, McShea, & Synovec, 2010; Machonis, Jones, Schaneberg, Kwik-Uribe, & Dowell, 2014; Van Durme, Ingels, & De Winne, 2016), which provide reliable and accurate descriptions of individual cocoa qualities. Recently, the use of novel methods based on on-line comprehensive two-dimensional liquid chromatography coupled to tandem mass spectrometry (LC $\times$ LC-MS/MS) have allowed the characterization of new secondary metabolites of cocoa beans (Toro-Uribe, Montero, López-Giraldo, Ibáñez, & Herrero, 2018). However, these methods are not recommended for routine raw material quality controls as they are destructive, require specialised personnel, sample preparation and expensive equipments. Thus simpler, faster and non-destructive techniques are required.

An alternative is infrared spectroscopy (IR is a fast non-destructive analytical tool that needs little samples preparation), which is useful for both qualitative and quantitative analyses of molecules. Finally, the application of chemometric techniques to IR data provides a powerful tool to develop methods capable of

classifying or characterising samples (Vergara-Barberán, Lerma-García, Herrero-Martínez, & Simó-Alfonso, 2015).

Recent studies into commercially cocoa derivates have shown the suitability of near IR (NIR) for the quantification of main constituents like fat, protein, moisture and carbohydrates (Veselá *et al.*, 2007). NIRS has also been shown appropriate to determine parameters like geographical origin (Teye, Huang, Dai, & Chen, 2013), fermentation quality (Hue *et al.*, 2014), fermentation index or pH (Sunoj, Igathinathane, & Visvanathan, 2016), and minor valuable components like organic acids (Krähmer *et al.*, 2015), caffeine, theobromine and epicatechin in unfermented and sun-dried beans (Álvarez, Pérez, Cros, Lares, & Assemat, 2012), procyanidins (Whitacre *et al.*, 2003) and adulterations with carob flour (Quelal-Vásconez, Pérez-Esteve, Arnau-Bonachera, Barat, & Talens, 2018) and cocoa shell (Quelal-Vásconez, Lerma-García, Pérez-Esteve, Arnau-Bonachera, Barat, & Talens, 2019). Despite these advances in cocoa characterisation by NIRS, as far as we know, the effect of cocoa powder processing on methylxanthine and flavanol content by this technique has not yet been studied.

In this context, the objective of this work is twofold. Firstly, to study the effect of cocoa powder processing on methylxanthines (theobromine and caffeine) and flavanols (catechin and epicatechin) contents in a large batch of samples (with different origins, processing parameters and alkalisation levels). Secondly, to evaluate the feasibility of reflectance NIRS combined with partial least square (PLS) to non-destructively predict the content of these compounds in cocoa powders.

2. Materials and methods

2.1 Reagents and samples

The employed reagents were: caffeine, theobromine, catechin and epicatequin (Sigma-Aldrich, St. Louis, Missouri, USA), acetonitrile (J.T. Baker, The Netherlands), methanol (Labkem, Barcelona, Spain) and acetic acid glacial (Sharlau,

Barcelona, Spain). Deionised water was obtained using an Aquinity deioniser (Membrapure GmbH, Berlin, Germany).

Samples comprised 86 cocoa powders provided by Olam Food Ingredients Company (Cheste, Spain) or purchased in different nationals and international markets, to assure variability in cocoa origin (Ivory Coast (n = 22), Ghana (n = 8), Indonesia (n = 12), Ecuador (n = 7), Peru (n = 4), or undeclared (n = 33), year of production (2017 (n = 67) or 2018 (n = 19)) and processing practices (natural or alkalised cocoa powders).

Before analysing, samples were characterised according to their extractable pH and colour. Extractable pH determination was performed by the procedure described in the ADM Cocoa Manual (2006). CIE L*a*b colour coordinates were obtained by measuring the reflection spectrum using a 10° observer and D65 illuminant (Minolta, CM 3600D, Tokyo, Japan). Extractable and colour data were used to classify cocoa powders into five categories: natural (NC) (pH 5.0–6.0), light-alkalised (LAC) (pH 6.0–7.2), medium-alkalised (MAC) (pH 7.2–7.6), strong-alkalised (SAC) (pH > 7.6) and black powders (BC) (pH > 7.6 and very low L values) (Miller *et al*., 2008).

2.2. Instrumentation and experimental conditions

2.2.1. Methylxanthines and flavanols extraction

In order to extract methylxanthines and flavanols from cocoa powders, the protocol of Lacueva *et al*. (2008) was adapted: 0.5 g of cocoa powder was weighed, suspended in 5 mL H_2O at 100ºC+20 mL methanol, mixed for 20 min at 36ºC by constant agitation, and centrifuged for 10 min at 10,000 rpm at room temperature. Finally, the obtained supernatants were filtered using 0.22 µm pore size PTFE filters (Scharlab, Barcelona, Spain). The obtained samples were immediately injected into the HPLC system or stored at -20ºC.

Article 3

2.2.2. HPLC determination of methylxanthines and flavanols

Analyte determination was performed in a liquid chromatograph model LaChrom Elite (Hitachi Ltd., Tokyo, Japan), equipped with an auto-sampler and a UV detector (models L-220 and L-2400, respectively). A 5-µm analytical column was used, model Liquid Purple C18 (250 x 4.6 mm i.d.) from Análisis Vínicos (Tomelloso, Spain). The mobile phase was prepared by mixing 2% aqueous acetic acid (phase A) and a ternary mixture composed of acetonitrile–H_2O–acetic acid in the 40:9:1 v/v/v proportion (phase B). The following gradient (adapted from Elwers *et al.*, 2009) was used: 0-8 min, 10% B; 8-20 min, 10-15% B; 20-35 min, 15-90% B; 35-50 min, 90% B, then back to the initial conditions. Other chromatographic conditions were: UV detection, 280 nm; column temperature, 40ºC; injection volume, 20 µL; flow rate, 1.2 mL min^{-1}.

The differences in analyte content with the various different alkalisation levels were established by a multifactorial ANOVA (95% confidence level of LSD; $p < 0.05$), constructed using Statgraphics Centurion XV from Manugistics Inc. (Rockville, MD, USA).

2.2.3. NIR spectra collection

The 86 cocoa powders were scanned in a FOSS NIR 5000 System spectrophotometer, equipped with a transport module (Silver Spring, MD, USA). Samples (ca. 5.0 g) were measured by filling a round sample cup (3.8 cm diameter x 1 cm-thick quartz windows) to preserve unvarying surface and thickness during spectral collection. For each sample, 32 sequential scans with 700 wavelengths were measured from 1100 and 2500 nm at 2-nm intervals. Samples were measured twice, and the mean spectra were employed for the statistical analysis.

2.3. Spectral data analysis

Calibration models to predict the methylxanthines (theobromine and caffeine) and flavanols (catechin and epicatechin) contents (previously established by HPLC) were constructed by PLS analysis of the spectral data using Unscrambler v10.4 software from the CAMO Software AS (OSLO, Norway). PLS is a multivariate calibration method that could be used to correlate NIR spectra data with chemical component contents. Spectral data were organised in a matrix containing the number of cocoa samples *(N = 86)* in rows and the X- and Y-variables in columns. The X-variables corresponded to the 700 wavelengths, while the Y-variables were the HPLC-determined analyte contents. For PLS model construction, the 86 samples were separated into two datasets: the 67 samples from, 2017 were employed to create and evaluate the PLS models by leave-one-out cross-validation, while the 19 samples from 2018 were used as an external validation set.

Before any spectral treatment, all spectra were used to construct a principal component analysis (PCA) model, which was employed to identify and remove defective spectral outliers using the Q residual values and the Hotelling T^2 with a 95% confidence limit (Bro & Smilde, 2014). Moreover, with the aim of detecting similarities among the samples, NIR spectra data were employed to build a clustering model by using the hierarchical complete-linkage method (HCL). This method aims to find similarities and patterns between samples, being this information represented in a dendrogram plot (Wajrock, Antille, Rytz, Pineau, & Hager, 2008).

Finally, PLS models were constructed with no spectral pretreatment (raw data) and with three different pre-treatments: 2nd derivative performed with the Savitzky-Golay smoothing algorithm (2nd derivative S-G) (Savitzky & Golay, 1951), orthogonal signal correction (OSC) (Wold, Antti, Lindgren, & Öhman, 1998), and their combination. The PLS models' accuracy was evaluated by: the required number of latent variables (LVs), the root mean square error of calibration (RMSEC) and cross-validation (RMSECV), and the coefficient of determination for calibration (R^2_C) and cross-validation (R^2_{CV}). The PLS models' predictive capability was judged

by considering: the coefficient of determination for prediction (R^2P), the root mean square error of prediction (RMSEP), the bias, the standard error of prediction (SEP) and the ratio of prediction deviation (RPD) (calculated as the ratio between the standard deviation of the reference values -training set- and RMSEP). Performance of the different models was considered good when the number of LVs, and RMSE and SE values were low, when R^2 value tended to unit, and when bias values were near to zero, for both the calibration and external prediction parameters (Quelal-Vásconez *et al*., 2019). Regarding RPD, a value between 2.0 and 2.5 makes approximate quantitative predictions possible, while values between 2.5 and 3.0, and above 3.0, indicates a prediction that could be considered good and excellent, respectively (Saeys, Mouazen, & Ramon, 2005).

3. Results and discussion

3.1 Alkalisation level evaluation

The 86 cocoa powders were physico-chemically characterised and divided into five categories according to their extractable pH values and colour (Miller *et al*., 2008). By contemplating these values, 23 cocoa samples were classified as natural, 19 as light-alkalised, 21 as medium-alkalised, 19 as strong-alkalised and 4 as black powders. Considering the linear correlation between alkalisation and pH (Pérez-Esteve, Lerma-García, Fuentes, Palomares, & Barat, 2016), this classification was taken into account to study the relationship between alkalisation intensity and the changes found in caffeine, theobromine, catechin and epicatechin contents.

3.2 HPLC determination of methylxanthines and flavanol contents in cocoa powders

By applying the experimental conditions included in Section 2.2.1, analyte peak identification was achieved by comparing the retention times of the sample peaks with the retention times of the standards. Four peaks were observed at retention times of 5.1, 9.1, 12.0 and 15.8 min, which respectively corresponded to theobromine,

150

catechin, caffeine and epicatechin. These compounds were quantified using external calibration curves. To construct them, six standard solutions at different concentrations within the ranges showed in Table 1 were prepared and injected. In all cases, R^2 above 0.9996 were observed. The other parameters in Table 1 were limit of detection (LOD) and limit of quantification (LOQ), which were estimated following the ICH guidelines (1996). The obtained values ranged between 0.005-0.1 and 0.017–0.33 mg L^{-1} for LODs and LOQs, respectively. These values were lower than those previously reported by others (Gottumukkala, Nadimpalli, Sukala, & Subbaraju, 2014; Risner, 2008; Srdjenovic, Djordjevic-Milic, Grujic, Injac, & Lepojevic, 2008). In order to assure that no matrix effect was observed in the quantification of analytes, standard addition calibration curves (considering the linearity ranges in Table 1) were constructed. The four curves provided R^2 above 0.9995 and similar slopes to the external calibration curves. Therefore, it was concluded that the external calibration curves were correctly used to quantify these analytes in cocoa powders.

Table 1 Linear ranges, determination coefficient, LOD and LOQ of the determination of methylxanthines and monomeric flavanols by HPLC.

Analyte	t_R (min)	Linear range (mg L^{-1})	R^2	LOD (mg L^{-1})	LOQ (mg L^{-1})
Theobromine	5.1	50-500	0.9998	0.005	0.017
Catechin	9.1	1-50	0.9996	0.1	0.33
Caffeine	12.0	5-250	0.9998	0.05	0.17
Epicatechin	15.8	0.5-100	0.9999	0.1	0.33

t_R: retention time; R^2: determination coefficient; LOD: limit of detection; LOQ: limit of quantification.

Finally, the 86 cocoa powders were subjected to HPLC (see Table 2 for the results). The most abundant alkaloid was theobromine, whose content ranged from 1.53±0.02 to 2.4±0.1 g/100g when considering all the samples. Statistical differences in theobromine content were found among alkalisation (p<0.05) as BC possessed lower contents than the other categories. The caffeine content ranged from 0.1503±0.0003 to 0.412±0.003 g/100g. As with theobromine, significant differences were observed among samples with different alkalisation degrees (p<0.05), and this behaviour has been reported by other authors (Brunetto *et al.* 2007; Li *et al.*, 2012).

Among flavanols, epicatechin was the most abundant analyte. The highest content (536.59±0.08 mg/100g) went to a NC sample labelled as Equator origin. Other natural samples exhibited an average content of ca. 160 mg/100 g. Contents statistically lowered as the alkalisation process became more intense, and reached average values of ca. 80, 33, 28 mg/100 g and with LODs in the LAC, MAC, SAC and BC, respectively. The same behaviour was found for catechin contents. In the NC samples, catechin content ranged from 15.2±0.5 to 167±1.2 mg/100g. The highest value went to another sample from Equator. In the other natural samples, the average value was ca. 80 mg/100 g.

Table 2. Theobromine, catechin, caffeine and epicatechin content (mean, minimum and maximum values and their standard error) found for the different types of cocoa powders employed in this study.

| Analyte	Alkalization degree*				
	Natural (pH 5-6)	Light (pH 6-7.2)	Medium (pH 7.2-7.6)	Strong (pH >7.6)	Black (pH >7.6, low L*)
Theobromine (g/100g)	2.15[a] (1.53±0.02- 2.36±0.02)	2.13[ab] (1.77±0.01– 2.393±0.002)	2.04[ab] (1.68±0.02- 2.4±0.1)	2.06[ab] (1.76±0.01– 2.33±0.09)	1.95[b] (1.768±0.001- 2.07±0.01)
Catechin (mg/100g)	78.21[a] (15.2±0.5- 167.5±1.2)	60.39[b] (13.3±0.2–132±1)	27.86[b] (4.9±0.2–65.6±0.5)	23.36[b] (4.53±0.07–46.3±0.5)	<LOD
Caffeine (g/100g)	0.29[a] (0.201±0.002– 0.412±0.003)	0.24[b] (0.175±0.001– 0.393±0.002)	0.21[bc] (0.157±0.002- 0.334±0.001)	0.20[c] (0.172±0.002– 0.255±0.001)	0.20[bc] (0.1503±0.0003- 0.302±0.002)
Epicatechin (mg/100g)	163.17[a] (11.5±0.2- 536.59±0.08)	80.24[b] (11.9±0.2–219±1)	33.27[bc] (8.0±0.3-111.00±0.05)	28.13[c] (10.5±0.2-54.9±0.2)	<LOD

Within rows, values bearing different letters are significantly different ($p < 0.05$). LOD = limit of detection

According to Table 2, no significant differences in catechin content were found in the LAC cocoa powders, but contents statistically decreased in the medium-alkalised powders, with undetected quantities in the black powders. Similar results have been found by other authors. In a study done with 11 cocoa powder samples, Lacueva *et al* (2008) reported values of 200 and 70 mg/100 g for epicatechin and catechin in natural powders, and 30 and 25 mg/100 g in alkalised powders, respectively. Similar contents were observed by Payne *et al*. (2010), who reported 223 and 88 mg/100 g (natural), 69 and 70 mg/100 g (light-alkalised), 26 and 36 mg/100 g (medium-alkalised) and 4 and 9 mg/100 g (strong-alkalised) for epicatechin and catechin, respectively.

3.3 Clustering analysis of the samples

The clustering analysis of the NIR spectra shows four main groups (see Fig. 1). First group comprised only one sample of BC, while the second one included the other BCs and ca. 16% of the SAC samples. The third group included most NCs, ca. 74% of LACs, ca. 20% of MACs and ca. 6% of SACs. Finally, the fourth group included the 81% of MACs, ca. 79% of SACs, 26% of LACs and the rest of NCs. In general, it can be observed that BC samples (minimum content of catechin and epicatechin detected) belonged to the first and second group and that the rest groups are mostly linked to strong (group 4) or mild alkalisation conditions (group 3). However, this grouping cannot be completely linked neither to the content or flavanols nor methylxanines. For instance, samples with the highest flavanols content (Ecuadorian samples –i.e. NC 67 or NC 52-) are not clearly separated from other natural or alkalised samples. Therefore, it could be concluded that in sample clustering, besides quantified analytes or alkalisation degrees, there are other chemical signals (i.e. proteins, sugars, volatiles, etc) that are affecting sample clustering.
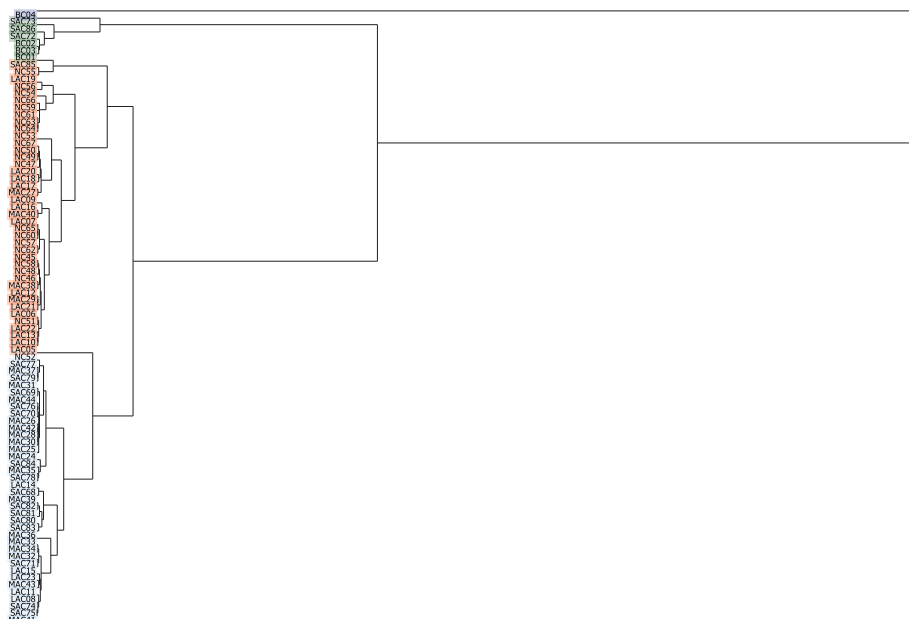
Figure 1. Dendogram of the clustering analysis performed with the NIR spectra of the entire sample set (n=86).

3.4. Prediction of theobromine, catechin, caffeine and epicatechin content by PLS

The raw spectra of the samples (between 1100 and 2500 nm) could contain both useful and irrelevant information; thus, it was pre-treated to enhance the final outcome. For all spectra, the first applied pre-treatment was the 2nd derivative S-G, followed by the OSC pre-treatment and their combination. Then, the optimal pre-treatment method to enhance the PLS models' performance was selected by taking into account the values included in Table 3. As can be observed in this table, the best results were observed after applying the 2nd derivative S-G+OSC for all models. Thus, the spectra obtained after applying this pre-treatment is shown in Fig. 2. As seen in this figure, signal peaks at wavelengths of 1728, 1764, 1884, 2312 and 2348 nm were evidenced. The region comprised between 1600 and 1800 nm predominantly corresponds to the first overtone region of carbohydrates (C–H bands) (Bázár et al., 2016), and the region comprised between 1700 and 2300 nm

normally attributes to the first overtones of C–H stretching associated with sugars (Osborne, Fearn, & Hindle, 1993, pp. 123–132).
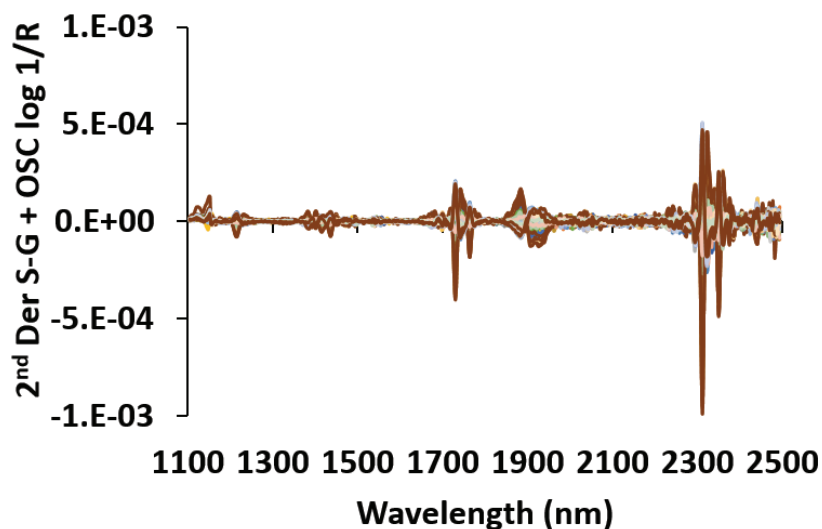
Figure 2. Spectra of the cocoa powders obtained from after applying 2nd derivative S-G + OSC pre-treatment.

Table 3 contains the results obtained for the PLS models constructed to predict the different analytes. As seen for the PLS models obtained with the optimal pre-treatment (marked in italics), all the PLS models provided satisfactory R^2_C, R^2_{CV} and R^2_P values within ranges 0.949-0.969, 0.802-0.889 and 0.813-0.819, respectively. For methylxanthine content predictions, the RMSEP values were 0.068 and 0.022 and the bias came very close to 0 (0.005 and 0.007) for theobromine and caffeine, respectively. With flavanols, the models provided RMSEP values of 8.160 and 7.430, and bias values of -1.440 and 1.034 for catechin and epicatechin, respectively. Notwithstanding, the RPD values for all the models were above 2.0, which are acceptable for quantitative predictions according to the literature (Saeys et al., 2005). The good fit between the different analytes content measured by HPLC and the contents predicted by the PLS models for the evaluation set samples is shown in Fig. 3.

Table 3. Results of the PLS models constructed for the prediction of theobromine, catechin, caffeine and epicatechin contents in cocoa powders.

Compound	Pre-treatment	#LV	Calibration		Cross-validation		Prediction			
			R^2_C	RMSEC	R^2_{CV}	RMSECV	R^2_P	RMSEP	bias	RPD
Theobromine (g/100g)	Non-pretreated	4	0.601	0.111	0.520	0.123	0.160	0.143	-0.027	1.22
	2nd Der. S-G	6	0.976	0.027	0.774	0.084	0.78	0.073	0.008	2.07
	OSC	4	0.923	0.048	0.870	0.064	0.793	0.209	0.027	1.38
	2nd Der. S-G + OSC	*3*	*0.969*	*0.030*	*0.889*	*0.058*	*0.819*	*0.068*	*0.005*	*2.13*
Catechin (mg/100g)	Non-pretreated	6	0.510	11.430	0.142	15.460	0.552	13.110	-1.463	1.25
	2nd Der. S-G	2	0.290	13.722	0.170	15.212	0.311	16.272	-1.962	1.01
	OSC	7	0.950	3.680	0.820	7.125	0.158	17.986	1.829	0.91
	2nd Der. S-G + OSC	*1*	*0.997*	*0.787*	*0.932*	*4.349*	*0.830*	*8.160*	*-1.440*	*2.01*
Caffeine (g/100g)	Non-pretreated	7	0.832	0.026	0.763	0.031	N.A	0.058	0.026	1.09
	2nd Der. S-G	5	0.941	0.015	0.763	0.031	0.775	0.025	0.008	2.55
	OSC	3	0.911	0.019	0.869	0.023	0.031	0.051	-0.028	1.22
	2nd Der. S-G + OSC	*1*	*0.940*	*0.015*	*0.802*	*0.028*	*0.813*	*0.022*	*0.007*	*2.81*
Epicatechin (mg/100g)	Non-pretreated	7	0.900	5.210	0.433	12.924	0.690	9.899	-4.690	1.58
	2nd Der. S-G	5	0.972	2.740	0.295	14.412	0.644	10.592	-3.024	1.57
	OSC	1	0.999	0.050	0.999	12.968	0.470	12.970	-0.480	1.29
	2nd Der. S-G + OSC	*6*	*0.999*	*0.262*	*0.988*	*1.810*	*0.824*	*7.430*	*-1.034*	*2.24*

2nd Der. S-G = Second derivative-Savitzky Golay; OSC = Orthogonal signal correction; #LV = latent variables; R^2_C = coefficient of determination for calibration; RMSEC = root mean square error of calibration; R^2_{CV} = coefficient of determination for cross-validation; RMSECV = root mean square error of cross-validation; R^2P = coefficient of determination for prediction; RMSEP = root mean square error of prediction; RPD = ratio of prediction deviation; N.A. = non available.

The good fit between the different analytes content measured by HPLC and the contents predicted by the PLS models for the evaluation set samples is shown in Figure 4.

Article 3
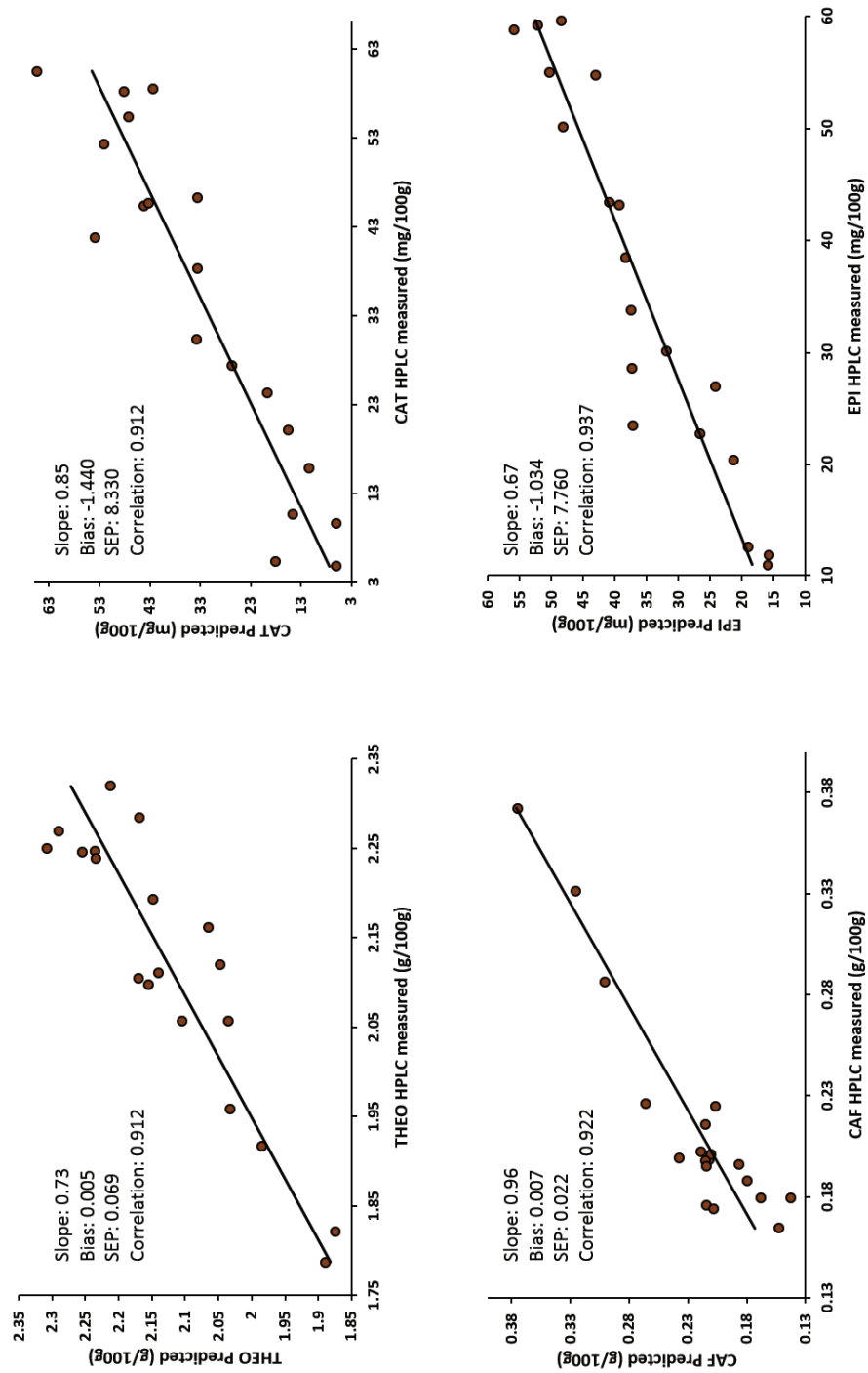
Figure 3. HPLC measured versus NIR predicted (a) theobromine (THEO), (b) catechin (CAT), (c) caffeine (CAF) and (d) epicatechin (EPI) contents by PLS in the prediction

Finally, the wavelengths with a better prediction capability according to the b vector profiles are shown in Fig. 4. The most important wavelengths related to variation in theobromine content (see Fig. 4a) are found at 1384, 1414, 1710, 1730, 1742, 1764, 1884, 1906, 1934, 2308, 2322 and 2350 nm, among others, and agree with those described by Álvarez *et al*. (2012). Among these wavelengths, 1414, 1730, 1764, 1906, 1934, 2308 and 2350 nm increased as theobromine content rose, while the rest lowered. According to literature (Veselá *et al*., 2007), these variations are characterised mainly by the stretching of H_2O of weakly bounded water, proteins and aromatics, the first overtone of stretching of CH of aromatics, the first overtone of the symmetric and asymmetric stretching vibration of CH_2, and the stretching and rocking vibrations of CH_2 of polysaccharides and fats.
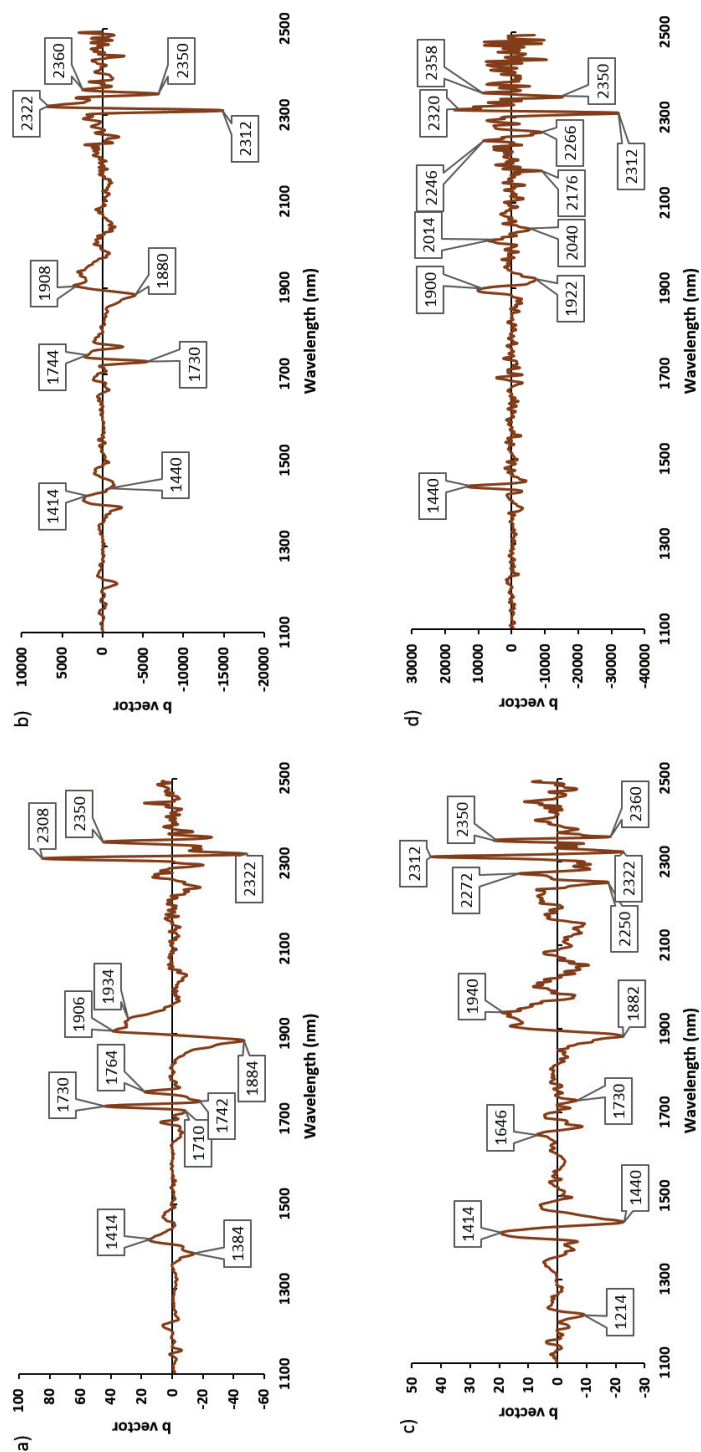
Figure 4. The B vector profiles of the PLS models constructed to predict (a) theobromine, (b) catechin, (c) caffeine and (d) epicatechin contents.

For catechin content predictions, the most important wavelengths were 1414, 1440, 1730, 1744, 1880, 1908, 2312, 2322, 2350 and 2360 nm (see Fig. 4b). Among them, 1414, 1744, 1908, 2322 and 2360 nm increased as catechin content rose, while the rest lowered. These variations may be associated with the stretching of H_2O of weakly bounded water, proteins and aromatics and also of non-bounded water, the asymmetric stretching and rocking of H_2O of very weakly bounded water fat, asymmetric stretching and rocking of CH_2 of polysaccharides and the stretching and rocking of CH and CC (Veselá *et al.*, 2007).

In the case of caffeine (see Fig. 4c), the most important wavelengths are located at 1414, 1646, 1940, 2272, 2312 and 2350 nm (which increase as the caffeine content raised) and the wavelengths at 1214, 1440, 1730, 1882, 2250, 2322, and 2360 nm (which decrease as the caffeine content raised). In this case, the wavelengths corresponded to the stretching of H_2O (non-bounded water), the first overtone of asymmetric stretching vibration of CH_2, proteins, combination of amides, the stretching of CH and CC and the stretching and rocking of CH_2 of fats (*Krähmer et al.*, 2015).

Finally, and as shown in Fig. 4d for epicatechin prediction, the most important wavelengths related to the variation of epicatechin are the 1440, 1900, 2014, 2246, 2320 and 2358 nm bands (which increase as the epicatechin content increases), and the 1922, 2040, 2176, 2266, 2312 and 2350 nm bands (which decrease as the epicatechin content raises). These bands could be assigned to the stretching of H_2O (non-bounded water), the asymmetric stretching and rocking of H_2O of very weakly bounded water fat, proteins and the stretching of CH and C=C (Veselá *et al.*, 2007). Similar NIR chemical vibrations has been found in the quantification of total phenols and carotenoids in blackberries (Toledo-Martín *et al.*, 2018), and in the determination of the fat, caffeine, theobromine of sun dried cocoa beans (Álvarez *et al.*, 2012).

Article 3

4. Conclusions

The HPLC determination of methylxanthines and flavanols of a large collection of cocoa powders provided not only better knowledge of the concentration variability in natural cocoas from different origins, but also the understanding of the effect that industrial alkalisation has on those contents. A wide fluctuation in flavanol content was observed for natural powders, which highlights the natural variability of unprocessed samples given their different origins. Moreover, the effect of cocoa alkalisation on the content of all analytes was evidenced. Despite such evidence, the many analysed samples indicated strong-alkalised powders with higher analyte contents than some natural cocoa powders. This reinforces the importance of measuring the content of these analytes during raw material selection and in all the industrial processing steps when functional products want to be launched on the market. In line with this, the possibility of predicting the content of these functional analytes by a fast, non-destructive and reliable methodology, such as NIRS, was confirmed. Despite the fact that the clustering analysis did not allowed a sample grouping according to the alkalisation degree or the flavanols and/or methylxantines content by applying PLS models, all analytes were satisfactorily predicted. Therefore, the present results bridge the information gap in the cocoa sector about the variability found in these functional compounds in commercial samples, and also propose a fast reliable methodology to establish the content of these important functional compounds for the cocoa industry.

162

Conflicts of interest

The authors declare that they have no conflict of interest.

References

ADM Cocoa Manual. (2006). The De Zaan® cocoa manual. Archer Daniels Midland Company BV. The Netherlands.

Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2008). Flavor Formation and Character in Cocoa and Chocolate: A Critical Review. *Critical Reviews in Food Science and Nutrition, 48*(9), 840–857.

Álvarez, C., Pérez, E., Cros, E., Lares, M., Assemat, S., Boulanger, R., & Davrieux, F. (2012). The use of near infrared spectroscopy to determine the fat, caffeine, theobromine and (–)-epicatechin contents in unfermented and sun-dried beans of Criollo cocoa. *Journal of Near Infrared Spectroscopy, 20*(2), 307-315.

Bázár, G., Romvári, R., Szabó, A., Somogyi, T., Éles, V., & Tsenkova, R. (2016). NIR detection of honey adulteration reveals differences in water spectral pattern. *Food Chemistry, 194*, 873–880.

Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods, 6*(9), 2812–2831.

Brunetto, M. del R., Gutiérrez, L., Delgado, Y., Gallignani, M., Zambrano, A., Gomez, A., & Romero, C. (2007). Determination of theobromine, theophylline and caffeine in cocoa samples by a high-performance liquid chromatographic method with on-line sample cleanup in a switching-column system. *Food Chemistry, 100*(2), 459-467.

Cádiz-Gurrea, M. L., Lozano-Sanchez, J., Contreras-Gámez, M., Legeai-Mallet, L., Fernández-Arroyo, S., & Segura-Carretero, A. (2014). Isolation, comprehensive characterization and antioxidant activities of Theobroma cacao extract. *Journal of Functional Foods, 10*, 485–498.

Elwers, S., Zambrano, A., Rohsius, C., & Lieberei, R. (2009). Differences

Article 3

between the content of phenolic compounds in Criollo, Forastero and Trinitario cocoa seed (Theobroma cacao L.). *European Food Research and Technology, 229*(6), 937–948.

Fayeulle, N., Vallverdu-Queralt, A., Meudec, E., Hue, C., Boulanger, R., Cheynier, V., & Sommerer, N. (2018). Characterization of new flavan-3-ol derivatives in fermented cocoa beans. *Food Chemistry, 259*, 207–212.

Franco R., Oñatibia-Astibia, A., & Martínez-Pinilla, E. (2013). Health benefits of methylxanthines in cacao and chocolate. *Nutrients, 5*(10), 4159–73.

Gottumukkala, R. V. S. S., Nadimpalli, N., Sukala, K., & Subbaraju, G. V. (2014). Determination of Catechin and Epicatechin Content in Chocolates by High-Performance Liquid Chromatography. *International Scholarly Research Notices, 2014*, 1–15.

Hue, C., Gunata, Z., Bergounhou, A., Assemat, S., Boulanger, R., Sauvage, F. X., & Davrieux, F. (2014). Near infrared spectroscopy as a new tool to determine cocoa fermentation levels through ammonia nitrogen quantification. *Food Chemistry, 148*, 240–245.

Humston, E. M., Knowles, J. D., McShea, A., & Synovec, R. E. (2010). Quantitative assessment of moisture damage for cacao bean quality using two-dimensional gas chromatography combined with time-of-flight mass spectrometry and chemometrics. *Journal of Chromatography A, 1217*(12), 1963–1970.

International Conference on Harmonization (ICH guidelines). (1996). Validation of analytical procedures: Text and methodology. Geneva: ICH-Q2.

Kongor, J. E., Hinneh, M., de Walle, D. Van, Afoakwa, E. O., Boeckx, P., & Dewettinck, K. (2016). Factors influencing quality variation in cocoa (Theobroma cacao) bean flavour profile - A review. *Food Research International, 82*, 44–52.

Krähmer, A., Engel, A., Kadow, D., Ali, N., Umaharan, P., Kroh, L. W., & Schulz, H. (2015). Fast and neat – Determination of biochemical quality parameters in cocoa using near infrared spectroscopy. *Food Chemistry, 181*, 152–159.

Lacueva, C. A., Monagas, M., Khan, N., Izquterdo-Pulido, M., Urpi-Sarda, M., Permanyer, J., & Lamuela-Raventós, R. M. (2008). Flavanol and flavonol contents

of cocoa powder products: Influence of the manufacturing process. *Journal of Agricultural and Food Chemistry*, *56*(9), 3111–3117.

Langer, S., Marshall, L. J., Day, A. J., & Morgan, M. R. (2011). Flavanols and methylxanthines in commercially available dark chocolate: a study of the correlation with nonfat cocoa solids. *Journal of agricultural and food chemistry*, *59*(15), 8435–8441.

Li, Y., Feng, Y., Zhu, S., Luo, C., Ma, J., & Zhong, F. (2012). The effect of alkalization on the bioactive and flavor related components in commercial cocoa powder. *Journal of Food Composition and Analysis*, *25*(1), 17-23.

Machonis, P., Jones, M., Schaneberg, B., Kwik-Uribe, C., & Dowell, D. (2014). Method for the determination of catechin and epicatechin enantiomers in cocoa-based ingredients and products by high-performance liquid chromatography: First action 2013.04. *Journal of AOAC International, 97*(2), 506–509.

Miller, K. B., Hurst, W. J., Payne, M. J., Stuart, D. A., Apgar, J., Sweigart, D. S., & Ou, B. (2008). Impact of alkalization on the antioxidant and flavanol content of commercial cocoa powders. *Journal of Agricultural and Food Chemistry, 56*(18), 8527–8533.

Oñatibia-Astibia, A., Franco, R., & Martínez-Pinilla, E. (2017). Health benefits of methylxanthines in neurodegenerative diseases. *Molecular Nutrition & Food Research*, *61*(6), 1600670.

Osborne, B. G., Fearn, T., & Hindle, P.H. (1993). Practical NIR spectroscopy with applications in food and beverage analysis. *Burnt Mill, Harlow, Essex, England: Longman Group,* 2nd ed., 123–132.

Payne, M. J., Hurst, W. J., Miller, K. B., Rank, C., & Stuart, D. A. (2010). Impact of fermentation, drying, roasting, and dutch processing on epicatechin and catechin content of cacao beans and cocoa ingredients. *Journal of Agricultural and Food Chemistry, 58*(19), 10518–10527.

Pérez-Esteve, É., Lerma-García, M. J., Fuentes, A., Palomares, C., & Barat, J. M. (2016). Control of undeclared flavoring of cocoa powders by the determination of vanillin and ethyl vanillin by HPLC. *Food Control, 67*, 171–176.

Article 3

Quelal-Vásconez, M. A., Lerma-García, M. J., Pérez-Esteve, É., Arnau-Bonachera, A., Barat, J. M., & Talens, P. (2019). Fast detection of cocoa shell in cocoa powders by near infrared spectroscopy and multivariate analysis. *Food Control, 99*, 68–72.

Quelal-Vásconez, M. A., Pérez-Esteve, É, Arnau-Bonachera, A., Barat, J. M., & Talens, P. (2018). Rapid fraud detection of cocoa powder with carob flour using near infrared spectroscopy. Food Control, *92*, 183–189.

Risner, C. H. (2008). Simultaneous Determination of Theobromine, (+)-Catechin, Caffeine, and (-)-Epicatechin in Standard Reference Material Baking Chocolate 2384, Cocoa, Cocoa Beans, and Cocoa Butter. *Journal of Chromatographic Science*, 46(10), 892–899.

Saeys, W., Mouazen, A.M., & Ramon, H. (2005). Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy, *Biosystems Engineering, 91*(4), 393-402.

Savitzky, A., & Golay, M. (1951). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, *40*(42).

Srdjenovic, B., Djordjevic-Milic, V., Grujic, N., Injac, R., & Lepojevic, Z. (2008). Simultaneous HPLC Determination of Caffeine, Theobromine, and Theophylline in Food, Drinks, and Herbal Products. *Journal of Chromatographic Science, 46*(2), 144–149.

Sunoj, S., Igathinathane, C., & Visvanathan, R. (2016). Nondestructive determination of cocoa bean quality using FT-NIR spectroscopy. *Computers and Electronics in Agriculture, 124*, 234–242.

Teye, E., Huang, X., Dai, H., & Chen, Q. (2013). Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *114*, 183-189.

Toledo-Martín, E. M., García-García, M. C., Font, R., Moreno-Rojas, J. M., Salinas-Navarro, M., Gómez, P., & del Río-Celestino, M. (2018). Quantification of total phenolic and carotenoid content in blackberries *(Rubus fructicosus L.)* using

166

near infrared spectroscopy (NIRS) and multivariate analysis. *Molecules, 23*(12), E3191.

Toro-Uribe, S., Montero, L., López-Giraldo, L., Ibáñez, E., & Herrero, M. (2018). Characterization of secondary metabolites from green cocoa beans using focusing-modulated comprehensive two-dimensional liquid chromatography coupled to tandem mass spectrometry. *Analytica Chimica Acta*, *1036*, 204-213.

Van Durme, J., Ingels, I., & De Winne, A. (2016). Inline roasting hyphenated with gas chromatography-mass spectrometry as an innovative approach for assessment of cocoa fermentation quality and aroma formation potential. *Food Chemistry, 205*, 66–72.

Vergara-Barberán, M., Lerma-García, M. J., Herrero-Martínez, J. M., & Simó-Alfonso, E. F. (2015). Cultivar discrimination of Spanish olives by using direct FTIR data combined with linear discriminant analysis. *European Journal of Lipid Science and Technology*, *117*(9), 1473–1479.

Veselá, A., Barros, A. S., Synytsya, A., Delgadillo, I., Čopíková, J., & Coimbra, M. A. (2007). Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder. *Analytica Chimica Acta, 601*(1), 77–86.

Wajrock, S., Antille, N., Rytz, A., Pineau, N., & Hager, C. (2008). Partitioning methods outperform hierarchical methods for clustering consumers in preference mapping. *Food Quality and Preference, 19*(7), 662-669.

Whitacre, E., Oliver, J., Van Den Broek, R., Engelsen, P., Kremers, B., Van Der Horst, B., Stewart, M., Jansen-Beuwink, A. (2003). Predictive Analysis of Cocoa Procyanidins Using Near-Infrared Spectroscopy Techniques. *Food Chemistry and Toxicology Predictive, 68*(9), 2618–2622.

Wold, S., Antti, H., Lindgren, F., & Öhman, J. (1998). Orthogonal signal correction of near-infrared spectra. *Chemometrics and Intelligent Systems*, *44*, 175–185.

4. GENERAL DISCUSSION

4. GENERAL DISCUSSION

The **first chapter** of this thesis shows the results aimed to solve a common problem in cocoa industries: the adulteration/mixing of cocoa powder with cheaper (carob flour) or valueless (cocoa shell) materials. Article (**I**) concerns the determination of cocoa shell percentage (undesired material) present in cocoa powders, which can be considered a parameter to be controlled during the shelling process as well and in processed products. The determination of cocoa shell percentage was done by means of the application of NIR and multivariate analysis. On the other hand, in the article (**II**), the detection of cocoa powder adulteration with carob flour was done using the same methodology.

To carry out the studies, a collection of 86 cocoa powders with different alkalization degrees and origins were gathered. To the mentioned samples, 6 carob flours of different roasting levels and 2 cocoa shells were added.

In the study dealing with the prediction of cocoa shell content by using NIR (article I), from the mentioned collection, all 20 natural cocoas (coming from different origins and obtained after different processings) part of them obtained in a company that guarantees the lowest cocoa shell content were randomly selected and mixed with the 2 cocoa shell samples providing a total of 72 mixtures. Cocoa shell was added to cocoa powder samples in 7 percentages which were 0, 2.5, 5, 7.5, 10, 20 and 40%. This maximum value was fixed since higher percentages would provide a product with physico-chemical characteristics (for example texture) different to cocoa.

After the spectra acquisition, differences among spectra of pure cocoa shell and pure cocoa powders were found. These differences allowed a separation of samples by means of a PCA analysis. After this, different pretreatments were applied to the spectra, being the EMSC +OSC pretreatment the one that led to better results for both, PLS and PLS-DA models construction. For both cases, the use of all spectra wavelengths was considered, jointly with the use of the most important wavelengths

selected according to the Variable Importance in Projection (VIP) scores, models being next compared. The best prediction performance was obtained with a PLS model constructed with using the 6 main wavelengths selected according to VIP scores and 1 latent variable, which provided R^2_P of 0.967, RMSEP of 2.43% and a RPD of 5.03.

On the other hand, the same selected wavelengths were used for the construction of the PLS-DA model aimed to classify samples into two categories: cocoa powders containing shell contents (based on fat-free dry matter) below 5% and comprised between 5 and 40% (since 5% is considered the acceptance limit in cocoa powders by the Codex Alimentarius). The best PLS-DA model provided sensitivity, specificity and NER values of 100, 85 and 92.5%, respectively, for the first category (below 5%), while for the second one (5-40%) the values reported were 85, 100 and 92.5%, respectively.

In the prediction of cocoa powder adulteration with carob flour (article II), 12 cocoas with proven authenticity with 4 alkalization degrees (3 cocoa for each alkalization degree) were mixed with the 6 carob. Carob flours were added to cocoa powder samples in percentages comprised between 0 and 40%. In this study, the Latin Hypercube Strategy (LHS) (Helton & Davis, 2003) was followed to perform a total of 216 mixtures.

After the spectra measurements, the first exploration of the results done with a PCA showed different groups of samples. The most variability was related to the alkalization degree of cocoa samples and to the differences among cocoas and carob flours. After this, different pretreatments were applied to the spectra, being the 2nd Der. S-G + OSC pretreatment the one that led to better results. For all the PLS models and the PLS-DA, the use of all spectra wavelengths was considered. Using these pretreated spectra, a PLS-DA was built in order to classify samples into three categories: pure cocoa powders, adulterated cocoa powders and carob flours. A good PLS-DA model was obtained with an accuracy value of 100% for the classification of the samples of the external validation set in the three categories. Finally, the best

prediction of the adulteration percentage was obtained by means of PLS using 1 latent variable, providing R^2_P, RMSEP of 3.2% and a RPD of 6.3.

Taking into account the results previously indicated for both qualitative and quantitative analysis of both articles, it can be concluded that NIR spectra combined with multivariate analysis is a good tool to assess in a fast and non-destructive way the presence of valueless (cocoa shell) or cheaper materials (carob flour) in cocoa powders.

The **second chapter** of this thesis deals with the fast analysis of some bioactive compounds present in cocoa powders (methylxanthines and flavanols). This chapter contains the article (**III**) in which the content of these bioactive compounds were determined by HPLC in a huge cocoa powder sample set (n=86), covering different origins, processing parameters and alkalization levels, being these content next predicted by using NIR and PLS regression.

By one hand, the variation of methylxanthines (theobromine and caffeine) and flavanols (catechin and epicatechin) content in a large cocoa powders sample set, was first established by HPLC. HPLC results showed on the one hand, for the methylxanthines values of theobromine in the range of 2.15 to 1.95 g/100g, caffeine values since 0.29 to 0.20 g/100g. On the other hand, for the flavanols values of catechin since 78.21 to 23.36 mg/100g, and for the epicatechin values in the range of 163.17 to 28.13 mg/100g. Concluding that the alkalization process led to a reduction of the content of all analytes, being this loss more evident in flavanols. The determination of these analytes in a huge set of samples has allowed not only a better knowledge of the concentration variability in natural cocoas coming from different origins, but also the understanding of the effect that industrial alkalization provokes in those contents.

Since HPLC is a destructive technique that requires long analysis times, the development of an alternative and fast analytical approach able to solve this problem is considered. In this regard, the use of NIR followed by PLS was successfully proposed. Concretely, all analytes were well-predicted, with a better prediction for

methylxanthines (R^2_p 0.882 for both analytes; RMSEP 0.020-0.061 g/100g, bias -0.027-0.006) than for flavanols (R^2_p 0.818-0.863; RMSEP 6.63-15.87 mg/100g, bias 1.942-3.056). Moreover, the models had values of RPD higher than 3 in all the prediction models, which means that those models are good.

5. CONCLUSIONS AND FUTURE PERSPECTIVES

CONCLUSIONS

This point summarizes the main conclusions of this thesis and draws out their implications for cocoa powder quality control. Five main general conclusions can be drawn from the results of this work:

1. A collection of 86 cocoa samples with different origins and alkalization degrees, 6 carob flours and 2 cocoa shells were gathered and strategically used in the studies of this thesis. Besides the mentioned samples, a battery of mixtures with known levels of cocoa and undesired materials (shell or carob flour) was constructed to be used during the correlation of a NIR spectrum with a specific content of both materials. This complete set of samples with different characteristics allowed to obtain robust models.

2. The fast detection of cocoa shell in cocoa powder is possible by means of NIR spectroscopy and multivariate analysis. The best results (R^2_P of 0.967 and a RMSEP of 2.43%) are found when the PLS model is built with the EMSC+OSC pretreated spectra, after the selection of the most important 6 wavelengths using VIP scores and 1 LV. Regarding the classification, the PLS-DA model allows the correct classification of 92.5% of the samples of a validation set in two categories, according to the Codex Alimentarius recommendations (<5% and >5% of cocoa shell content).

3. NIR spectroscopy combined with multivariate analysis is a rapid and non-destructive method to determine quantitative and qualitative the adulteration of cocoa powder with carob flour without any sample preparation. The PLS-DA model constructed was able to correctly classify the 100% of the validation set samples. The best prediction

177

model was obtained using PLS regression of the full spectra (1100-2500 nm) after preprocessing with 2nd Der. S-G+OSC with a R^2_P of 0.974 and a RMSEP of 3.2%.

4. The determination of methylxanthines and flavanols content of cocoa powders by means of HPLC allow to observe a wide range of values for each of the analytes. These values were partially related to the alkalization level, since in average, higher values were found in natural powders, as well as lower values were found in highly alkalized ones. Nevertheless, it was observed that inside a category (especially in natural cocoas), values fluctuated a lot among samples, indicating the natural variability of unprocessed samples probably due to their different origins, varieties and harvesting conditions. Furthermore, it was also found that some alkalized samples presented higher flavanols content than certain natural ones. This highlights the importance of measuring the content of these analytes during raw material selection and during the industrial processing steps to obtain functional products.

5. NIR spectroscopy can be used as an alternative to the HPLC, as a fast and non-destructive method in the routine determination of catechin, epicatechin, caffeine and theobromine in the cocoa powder industry. Good performance of the PLS models were obtained for the determination of methylxanthines (R^2_P of 0.882 for both analytes; RMSEP of 0.020 and 0.061%, BIAS of 0.006 and 0.027 for caffeine and theobromine, respectively) and flavanols (R^2_P of 0.863 and 0.818; RMSEP of 6.63 and 15.87%, BIAS of 1.942 and 3.056 for catechin and epicatechin, respectively).
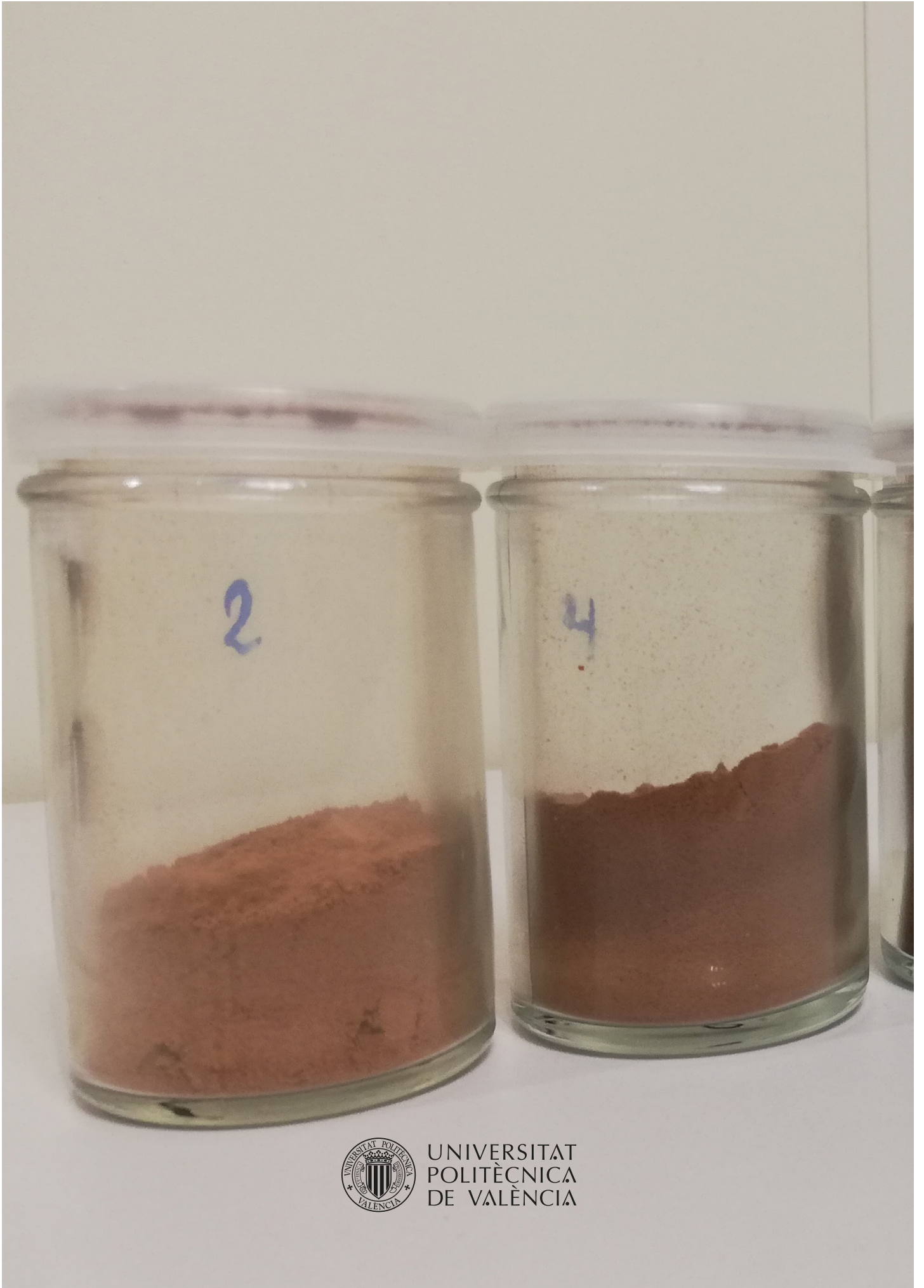
FUTURE PERSPECTIVES

The prediction and classification models built in this study were constructed using samples that covered different origins and alkalization degrees. Nevertheless, more natural variability could be considered. This natural variability in the cocoa intrinsic quality parameters is dependent also of environmental conditions, soil nutrients, fermentation practices, etc. which have not been considered in this study. This means that recalibration with new samples or with a higher number of samples might certainly improve the prediction and the discrimination capability of the models.

In comparison with models obtained in articles I and II, RPD values are lower than those of the article III. To obtain more accurate models, a strategic calibration of models according to the alkalization levels could allow more accurate models, over all in the case of flavanols and methylxanthines content prediction from NIR spectra. Another strategy to increase the accuracy of the models could be the combined use of Near Infrared with Medium Infrared spectroscopy.

The possibility of detection of adulterants not only offers a final product with higher quality but also ensures its safety. Thus, this technique could be proposed to the cocoa sector as a reliable method to control products authenticity.

The understanding of alkalization process by means of NIR spectroscopy can give a wide insight of how to produce products with higher functional quality, selecting those technological parameters that avoid big losses of functional compounds. On the other hand, the fast determination of these compounds might allow the industry to quantify their content in each batch of product and include this information in both, product technical sheet and product label.